AD-A038 356 AIR FORCE WEAPONS LAB KIRTLAND AFB N MEX F/G 12/1
DATA ANALYSIS FOR STOCHASTIC PROCESSES.(U)
FEB 77 J E NEGRO
UNCLASSIFIED AFWL-TR-76-193 NL

1 OF 2
AD
A038356

AFWL-TR-76-193

AFWL-TR-
76-193

DATA ANALYSIS FOR STOCHASTIC PROCESSES

February 1977

Final Report

Approved for public release; distribution unlimited.

AIR FORCE WEAPONS LABORATORY
Air Force Systems Command
Kirtland Air Force Base, NM 87117

This final report was prepared by the Air Force Weapons Laboratory, Kirtland Air Force Base, New Mexico, under Job Order 317J0798. Dr. Merritt (LRO) was the Laboratory Project Officer-in-Charge.

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

This report has been reviewed by the Information Office (OI) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

Paul H. Merritt

PAUL H. MERRITT
Project Officer

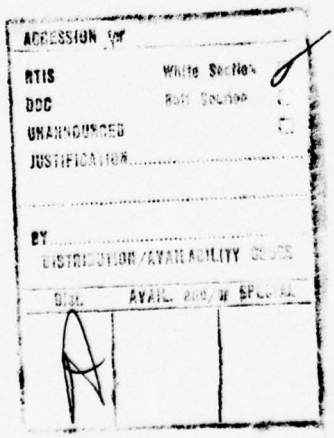
FOR THE COMMANDER

Lawrence Sher

LAWRENCE SHER
Chief, ALL Beam Control Branch

Demos T. Kyrazis

DEMOS T. KYRAZIS
Colonel, USAF
Chief, Laser Development Division



DO NOT RETURN THIS COPY. RETAIN OR DESTORY.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: AFWL-TR-76-193	2. GOVT ACCESSION NO. / 3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): DATA ANALYSIS FOR STOCHASTIC PROCESSES.	5. TYPE OF REPORT & PERIOD COVERED: Final Report.
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): James E. Negro, Captain, USAF	8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Air Force Weapons Laboratory (LRO), Kirtland Air Force Base, NM 87117	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 63605F, 317J0798
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Weapons Laboratory, Kirtland Air Force Base, NM 87117	12. REPORT DATE: Feb[illegible]y 1977 / 13. NUMBER OF PAGES: 156
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): 159 p.	15. SECURITY CLASS. (of this report): Unclassified / 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report): Approved for public release; distribution unlimited. 3173	17 07
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)	
18. SUPPLEMENTARY NOTES	
19. KEY WORDS (Continue on reverse side if necessary and identify by block number): Data Analysis; Stochastic Process; SIGANAL; Coherence; Confidence Intervals; Power Spectral Density; Random Transfer Function Estimates	

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report is tailored for use as a companion document to SIGANAL: A MODULAR SIGNAL ANALYSIS PROGRAM, by Ramon A. Tenario, AFWL-TR-75-165, which is a comprehensive package of software routines for signal analysis applications. The purpose of this report is to yield understanding and interpretation of practical stochastic signal analysis procedures and results. Basic mathematical foundations for stochastic processes are presented and then extended to the calculations of Power Spectral Densities, cross-

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

213150

(Blk 20)

spectral densities, coherence functions, and transfer functions. Pratical aspects of data processing such as sampling, finite materials; and confidence bounds are explained. A simple numerical example is developed to demonstrate most of the analytical tools developed in the report.

PREFACE

This document is designed as a companion document to AFWL-TR-75-165, SIGANAL: A MODULAR SIGNAL ANALYSIS PROGRAM developed by Mr Ramon A. Tenorio, Computational Services Division, Air Force Weapons Laboratory. Mr Tenorio's programs greatly streamlined and consequently simplified hardware data analysis procedures at the Laboratory. This software immeasurably increased the capability for application of sophisticated analysis procedures to our data analysis. Aided by this new capability and strongly encouraged by Mr Tenorio, the author set himself to the task of demonstrating the utility and interpretation of modern random data analysis capability afforded by the new software.

My thanks to Tony Tenorio for his strong encouragement and his uniquely organized, efficiently and easily used analysis package. His efforts in programming several of the specialized routines used in the simulation analysis is gratefully acknowledged.

Many hours of interesting and enligntening discussions with Dr Paul Merritt have clarified my understanding of signal coherence analysis. I thank him for his insight.

CONTENTS

CONTENTS (Continued)

ILLUSTRATIONS

ILLUSTRATIONS (Continued)

ILLUSTRATIONS (Continued)

TABLES

SECTION I

INTRODUCTION AND GUIDE

This guide is especially tailored for use as a companion document to SIGANAL: A Modular Signal Analysis Program (reference 1). SIGANAL is a user's guide to a comprehensive package of software routines for signal analysis applications. These routines, coded in FORTRAN IV, process a user specified sequence of data manipulation routines which operate on digitized data. Presently coded routines facilitate calculation of signal statistics including means, standard deviations, signal value probability density and distribution functions, Fourier transforms (discrete), power and cross-spectral densities, coherence functions and linear system transfer functions. This guide demonstrates the usefulness to which these routines can be applied in extracting pertinent signal information.

1 PURPOSE OF SIGNAL ANALYSIS

Signals embody and represent characteristics of the systems which generate them. Signals generated by sensors, for example, convey basic information about the process being measured. The signal information, conveyed typically as an electric voltage or current, represents some physical quantity such as position, velocity, acceleration, force, pressure, temperature, luminosity, brightness, field strength, fluid flow, volume, weight, length, density, charge, or any of innumerable other quantities. In some cases, the analyst is concerned only with static measurements wherein a single number adequately represents the desired quantity. That the gravitational acceleration on the earth's surface is 9.81 meters/sec^2 is sufficiently accurate for almost everyone except the experimental and the gravity wave physicists. We shall not be concerned with these constant signals. We are interested in fluctuating signals representing time-varying and dynamic system variables. Fourier analysis and other transform techniques have been develcped to represent these signals in a frequency domain basis particularly suitable for calculating the response of linear dynamics systems or analyzing requirements of communication channels. Techniques of stochastic signal analysis reviewed in this guide are suitable for these purposes; however, their true utility resides in interpretation of randomly fluctuating signals which arise in phenomena where future events are not deterministically predictable from past events. These signals occur in statistical mechanics, quantum mechanics, turbulence and noise theories. This guide defines parametrizations of stochastic process and descriptive parameters which may be measured or estimated.

We adopt the philosophy that power-spectral densities (PSD's) are the most convenient format for characterizing stochastic processes. We emphasize this approach throughout. It has evolved from our experience that linear dynamic systems described by differential equations have input-output properties that are most readily characterized and viewed as frequency domain transfer functions rather than convolution integrals. Frequency domain characterizations of both system forcing functions and system transfer functions provide ready visualization of output signal characteristics (in the frequency domain). By our emphasis on the power-spectral density, we devote little attention to discussion of autocorrelation functions other than to define them and to derive their relationship with power spectra. Our preference is not universal.

2 USER'S GUIDE TO SIGNAL ANALYSIS

Concepts essential to the understanding and interpretation of practical stochastic signal analysis procedures and results are presented in Sections II and III. Section IV concludes these developments with the presentation, interpretation and comment on the use of SIGANAL routines to analyze random data synthesized by a numerically simulated dynamical system excited by random disturbance phenomena.

Basic mathematical foundations defining and relating properties of stochastic processes are presented in Section II. Introductory probability theory is briefly developed and expanded to define stochastic processes. Probability concepts are generalized to stochastic process characterizations by autocorrelation functions and power spectral densities. Attention is restricted to ergodic processes which most practical stochastic processes approximate. The mathematical basis for calculating PSD's, cross-spectral densities, coherence functions, and transfer function estimates is developed.

In Section III, the practical aspects of calculating stochastic process characterizations are reviewed. These aspects include data sampling, finite duration observation intervals, and confidence bounds. The Fast Fourier Transform technique for efficiently calculating signal spectra is introduced and specific calculation formulas presented. These topics review technique as implemented in the SIGANAL code, and accent the proper application of analysis procedures and interpretation.

Section IV is devoted to the development of a simple numerical signal analysis problem which exemplifies the basic interpretation principles and augments our intuition as to what to expect in the way of results. A first order control

system excited by error sensor noise and an exponentially correlated disturbance process is simulated. Theoretical power spectra and coherence functions are calculated from the known system transfer functions and the white noise disturbance processes. These "expected" results are compared to the quantities calculated with the signal analysis routines.

Readers having no prior familiarity with data analysis should proceed sequentially from Section II through Section IV and should augment his reading with material in the references. Those more knowledgeable with signal analysis procedures should scan Section II to familiarize themselves with notation (see also the nomenclature) and then may study the numerical example, referring back to Section III as required for reference.

SECTION II

PROBABILITY AND STOCHASTIC PROCESSES

1 INTRODUCTION

A detailed exposition of probability, stochastic processes, and random data analysis as presented in the references (2, 3, 4) is beyond the scope and intent of this guide. We do present the basic definitions and concepts required for an introductory understanding of these subjects, particularly in their application toward analyzing random effects observed in measurements. First, the notions of random effects are introduced to explain phenomena which deterministic system concepts cannot predict. So motivated, we characterize stochastic processes and how they interact with our systems.

2 BASIC CONCEPTS: RANDOMNESS

The basic notion of random data or random signals stems from our inability to deterministically characterize signals or phenomena we observe. To deterministically characterize a signal, we mean that we can specify explicity signal values as a function of time in units of voltage, amperage, pressure or some other appropriate unit. For example, we know that a voltage V of 10 volts impressed upon a 1 μf capacitor in parallel with a 1 MΩ resistor will decay as $V(t) = 10\ e^{-t}$ once the voltage source is removed. The voltage has been expressed as a well-known mathematical function of time. Another important aspect of the example is that we obtain the same result whenever the experiment is repeated.

Not all experiments are precisely repeatable. The radio interference from an automotive ignition system does not duplicate with precisely the same output radio signal. Other experiments involving fluid turbulence, electronic noise and molecular or atomic collisions, for example, result in phenomena which cannot be precisely predicted. These phenomena are random because they cannot be precisely specified. In subsequent sections, statistical characterizations of these processes shall be defined. These characterizations shall serve as a basis for describing the response of systems having such random excitations.

Development of analysis definitions and techniques is required for an understanding of the response of dynamic systems to random excitations. Application of these techniques is often essential for proper system design and system performance analysis. Just as a system designer can compute precisely the control system response to a deterministic input (e.g., step function, sinusoid,

exponential) we wish to be able to compute the response, on the average, to random inputs. Further, we must know the average variation by which one particular response may differ from the average response. These data allow specification of system sensors and actuators so that they will not be saturated by the noises and disturbances affecting the system. Statistical characterizations of random disturbances and their effect on dynamical systems also allow the designer to compute system accuracy or performance bounds for those cases in which random disturbances are the fundamental system limitations. With this motivation, we begin with a discussion of elementary probability theory. A comprehensive treatment is available in Papoulis (reference 4).

An experiment whose outcome cannot be predicted is said to be random. The result of a coin toss is therefore random since we cannot predict in advance whether the result will be HEADS or TAILS. There are some important aspects of this experiment, however, that can be described. For a fair coin, we would expect that the result HEADS is just as likely as the result TAILS. Intuitively then, we would expect that if the coin were tossed N times, HEADS would come up approximately N/2 times (i.e., $N_H \approx N/2$) and TAILS would come up $N_T \approx N/2$ times. The ratio N_H/N is called the relative frequency of the outcome HEADS in the coin toss experiment. We define the probability of the event HEADS as its relative frequency in the limit as the number of coin tosses approaches infinity. Notationally,

$$P(\{\text{HEADS}\}) = \lim_{N \to \infty} N_H/N \qquad (2\text{-}1)$$

Certain properties of the coin toss experiment are common to all random experiments. These properties are the basis of probability theory. Each experiment has a set of mutually exclusive results or outcomes. To each outcome we assign a positive number (possibly zero) representing the probability (relative frequency) that the outcome will occur if the experiment is repeated. Sets of outcome we call events and to each event we assign a probability equal to the sum of the component outcome probabilities. The event containing all possible outcomes has probability P = 1. These concepts are fully developed in reference 4, Chapter 2.

3 RANDOM VARIABLES

To each outcome of an experiment we may also assign an arbitrary number called a random number. The functional relationship between the number x and the events ζ is denoted $x(\zeta)$ and the function x is called a random variable, provided that it satisfies certain general conditions. Basically these conditions entail the requirement that the set $\{x \leq y\}$ corresponds to a set of experimental outcomes (i.e., an event). In our coin toss experiment, for example, we arbitrarily define a random variable as the rule which assigns value 1 to the outcome HEADS and value 0 to the outcome TAILS. We could have just as easily assigned the value 1/2 to HEADS and 1/2 to TAILS, as the particular values assigned are not of importance. A key concept of the random variable is that there exists a "one-to-one" correspondence between experimental events and sets of random numbers. Thus $\{x = 1\}$ corresponds to the events HEADS and $\{x = 0\}$, or more generally $\{x < 1\}$, corresponds to the event TAILS. In each case the random number associated with each outcome is contained in the given set of x values.

Since each set of random numbers is associated with an event of the underlying probability experiment, we associate with each set the properties of the underlying events. Most importantly, we associate with each set of random numbers the probability P of the associated event. The sets $\{x \leq y\}$ play a particularly important role. Any set of random variables associated with an event can be expressed in terms of the basic sets $\{x \leq y\}$ related by set operators union, intersection and complement. Properties of general sets are then obtained in terms of properties of the basic component sets. Important properties of these random variable sets are presented.

The probability distribution function is defined as the probability of the event $\{x \leq y\}$ and written

$$F(y) = P(\{x \leq y\}) \tag{2-2}$$

The probability distribution function for the random variable defined on the coin tossing experiment is plotted in figure 1.

The probability density function, pdf, is defined as the derivative of the probability distribution function. The density function for the coin toss experiment is plotted in figure 2. Notationally,

$$f(y) = \frac{d}{dy} F(y) \tag{2-3}$$

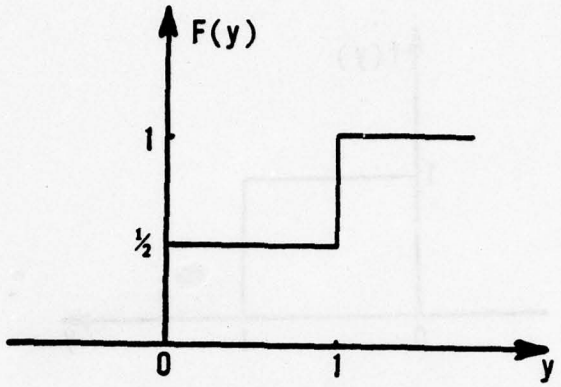


Figure 1. Coin Toss Experiment Probability Distribution Function.

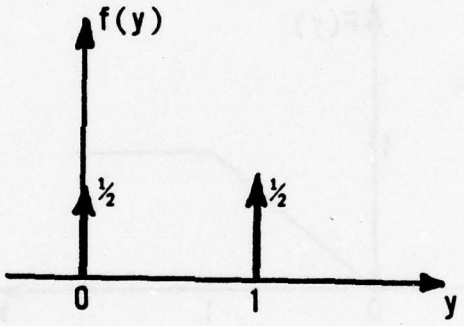


Figure 2. Coin Toss Experiment Probability Density Function.

The random variable associated with the coin tossing experiment is a discrete random variable since all events correspond to at most a countable number of specific random variable values. Discrete random variables have pdf's which are collections of impulse functions.

More generally an experiment may have a continuum of possible outcomes. The probability density function for these random variables is a continuous function except at possibly a countable number of points. The pdf for a uniformly distributed random variable is illustrated in figure 3. The corresponding probability distribution function is plotted in figure 4. This random variable is uniformly distributed since each value is equally likely to occur. Suppose we wish to know the probability that $\{y_1 < x \leq y_2\}$. By definition we know that

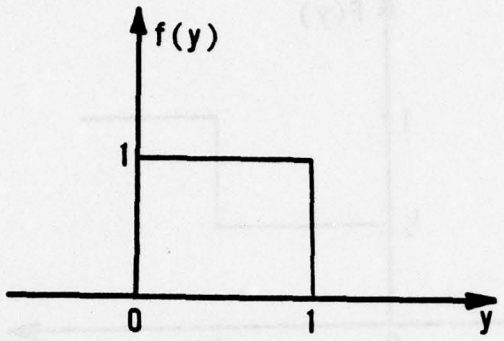


Figure 3. Uniformly Distributed Random Variable - pdf.

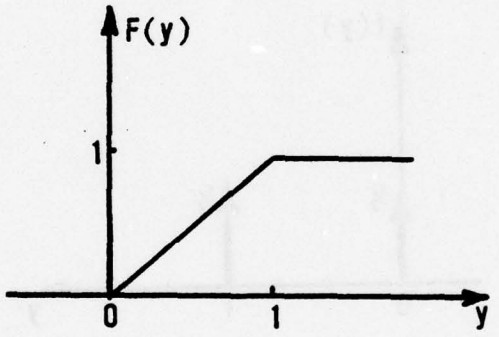


Figure 4. Uniformly Distributed Random Variable - Probability Distribution Function.

$$F(y_1) = P\ (\{x \leq y_1\}) \tag{2-4}$$

$$F(y_2) = P\ (\{x \leq y_2\}) \tag{2-5}$$

Because the event $\{x \leq y_2\}$ includes the event $\{x \leq y_1\}$ and since $\{y_1 < x \leq y_2\}$ equals $\{x \leq y_2\}$ less $\{x \leq y_1\}$, it follows that

$$P(\{y_1 < x \leq y_2\}) = P(\{x \leq y_2\}) - P(\{x \leq y_1\})$$

$$= F(y_2) - F(y_1) \tag{2-6}$$

By integration of equation (2-3) we have that

$$F(y) = \int_{-\infty}^{y} f(w)dw \tag{2-7}$$

Integral properties and equations (2-6) and (2-7) allow us to express equation (2-6) as

$$P(\{y_1 \prec x \preceq y_2\}) = \int_{y_1}^{y_2} f(w)dw \tag{2-8}$$

and for the case that $y_2 = y_1 + \Delta y$, Δy being a differential quantity, we have

$$P(\{y_1 \prec x \preceq y_1 + \Delta y\}) = f(y_1)\ \Delta y \tag{2-9}$$

Thus f(y) represents a differential probability since it can also be defined as the limit

$$f(y_1) = \lim_{\Delta y \to 0} \frac{P(\{y_1 < x < y_1 + \Delta y\})}{\Delta y} \tag{2-10}$$

The properties of the probability distribution and density functions expressed in equations (2-4) through (2-10) are true in general. Additional properties are presented in reference 3, Chapter 2, and reference 4, Chapter 4. For the uniform density, f(y) is a constant for $0 < y < 1$ so that by equation (2-10) the probability of any differential interval of random numbers is the same. That is, each differential interval is equally likely or has probability zero if the interval falls outside [0,1].

The normal, chi-square, binominal, beta, F-distribution and student's t-distribution are among those encountered in practice. We limit our attention to the normal and the chi-square distributions. These distributions play an important role in random data analysis. The normal distribution is particularly important in random variable theory since any random variable which is the sum of K identically distributed independent random variables has a density function that

approaches a normal distribution as K increases. This result is guaranteed by the central limit theorem (reference 5). The distribution and density functions for a normally distributed random variable are plotted in figures 5 and 6, respectively. The normal density function is written:

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-1/2\left(\frac{y-\mu}{\sigma}\right)^2\right] \qquad (2\text{-}11)$$

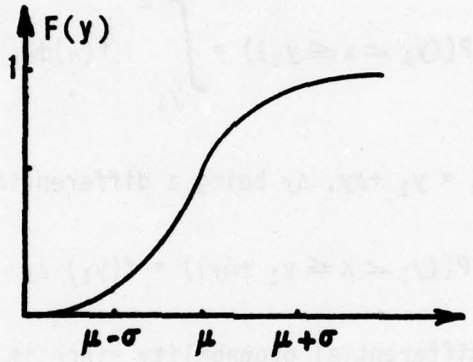


Figure 5. Normal Random Variable Probability Distribution Function.

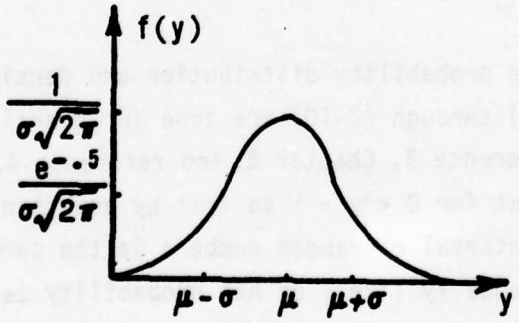


Figure 6. Normal Random Variable Probability Density Function.

The density function is completely specified by the parameters μ and σ. The plausibility of the central limit theorem is demonstrated in figure 7 in which density functions for the function W_K equal to the normalized sum of uniformly distributed independent random variables are plotted.

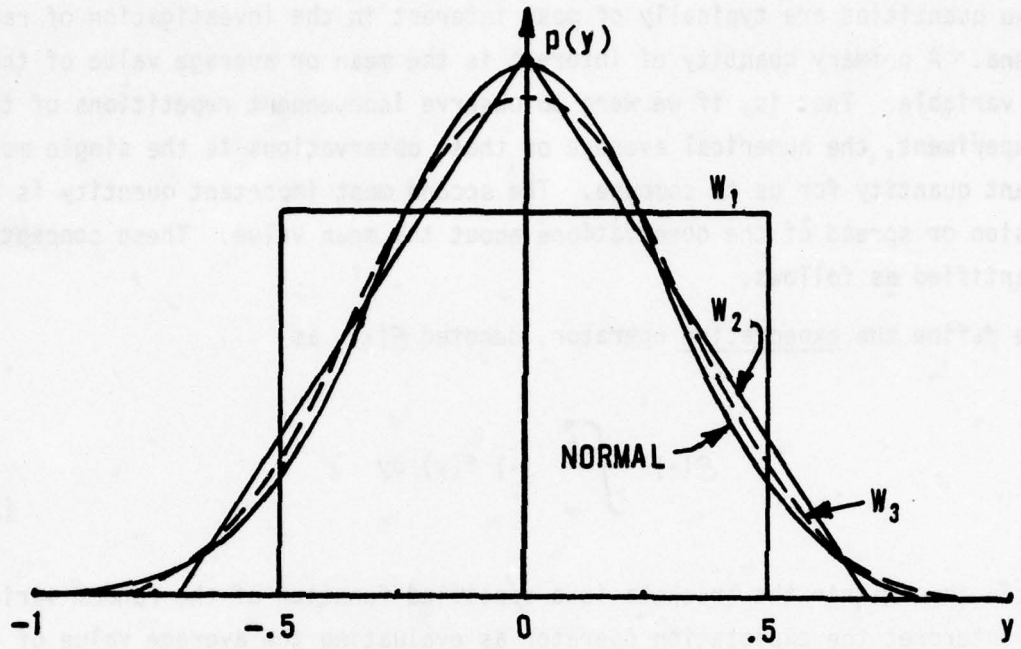


Figure 7. Demonstration of the Central Limit Theorem.

$$W_K = \frac{1}{\sqrt{K}} \sum_{i=1}^{K} x_i \qquad (2\text{-}12)$$

Also plotted in figure 7 is a normal density function. Observe that the density function for W_K approaches the normal density as K increases.

4 MEANS AND MOMENTS

Except for a few special cases, it is usually inconvenient or mathematically cumbersome to characterize random variables by specifying their probability density or probability distribution functions. These functions can be difficult to specify whenever they are not expressible in terms of known or tabulated functions as is done in equation (2-11). The pdf totally characterizes the random variable. An alternative approach to completely characterizing the random variable through its probability density function is to examine exactly what properties of the random variable we are interested in. Then we partially characterize the random variable in terms of parameterizations of these useful properties. This approach is exemplified through the following definitions and illustrative examples.

Two quantities are typically of most interest in the investigation of random phenomena. A primary quantity of interest is the mean or average value of the random variable. That is, if we were to observe independent repetitions of the same experiment, the numerical average of these observations is the single most important quantity for us to compute. The second most important quantity is the dispersion or spread of the observations about the mean value. These concepts are quantified as follows.

We define the expectation operator, denoted $\mathcal{E}[\cdot]$, as

$$\mathcal{E}[\cdot] = \int_{-\infty}^{\infty} [\cdot]\, f(y)\, dy \tag{2-13}$$

where the term within the brackets is a specified function of the random variable y. We interpret the expectation operator as evaluating the average value of the function upon which it operates. The average, of course, is with respect to the random variable and the underlying probability space upon which it is defined. Let us examine this important concept from another point of view. The average value of any quantity is simply the sum (integral) over all possible values that the quantity may assume of a product formed as a value multiplied by the probability that the particular value is assumed. Notationally

$$\text{Average } (g) = \int_{-\infty}^{\infty} g \cdot P(\{g < x \leq g + dg\}) \tag{2-14}$$

Application of equation (2-10) allows us to re-express equation (2-14) as

$$\text{Average } (g) = \int_{-\infty}^{\infty} g\, f(g)\, dg \tag{2-15}$$

We now realize that equation (2-13) is simply a variation of equation (2-15). The equations differ in the following respect. The bracketed quantity in equation (2-13) is a function, say g(y), of y. Particular values of g may be obtained by greatly different values of y. Rather than combining the probabilities of any y

which will yield the specified value g to obtain an equation of the form equation (2-15), we recognize that we can evaluate the probability density for each y separately and simply integrate over all possible y values to obtain equation (2-13). We illustrate this with an example.

EXAMPLE 2-1

Suppose we wish to calculate the average of the absolute value of the random variable y which is uniformly distributed in the interval [-1/2,1/2]. We have that:

$$f(y) = \begin{cases} 1, & -1/2 \leq y \leq 1/2 \\ 0, & \text{otherwise} \end{cases}$$

$$g(y) = |y|$$

By direct application of equation (2-13),

$$\mathcal{E}[|y|] = \int_{-\infty}^{\infty} |y| \, f(y) \, dy$$

$$= \int_{-1/2}^{1/2} |y| \, dy = 1/4$$

A pdf for g is obtained by first calculating the probability of the random variable g.

$$P(\{x < g \leq x + dx\}) = \begin{cases} P(\{x < y < x + dx\}) + P(\{x < -y \leq x + dx\}), & x > 0 \\ 0, \text{ otherwise} \end{cases}$$

Then by application of equation (2-10) we compute that:

$$f(g) = 2 \, f(y) \, , \; g \geq 0$$

$$0 \qquad , \text{otherwise}$$

Finally, evaluation of equation (2-15) gives

$$\text{AVERAGE } (g) = \int_{-\infty}^{\infty} g f(g)\, dg$$

$$= \int_{0}^{1/2} 2g\, dg = 1/4$$

We have obtained the same result by application of equation (2-13) or equation (2-15). Often in practice we shall find that equation (2-13) is easier to evaluate, generally because f(g) may not be readily expressed or as readily calculated as it is in this example.

The expectation operator provides a basis for defining two sets of quantities. The n^{th} moment of the random variable y is defined as

$$\mathcal{E}\left[y^n\right] = \int_{-\infty}^{\infty} y^n f(y)\, dy \tag{2-16}$$

The first moment is the mean or average value of the process and is denoted by μ_y. The nth central moment is defined as

$$\mathcal{E}\left[(y-\mu_y)^n\right] = \int_{-\infty}^{\infty} (y-\mu_y)^n f(y) dy \tag{2-17}$$

The first central moment is always equal to zero. It is of no interest to us. The second central moment is called the variance of the distribution and is denoted by σ_y^2. The square root of the variance is called the standard deviation of the process. Both the variance and the standard deviation characterize the dispersion of the distribution about its mean value. The larger the variance or standard deviation, the more probable it is that we will observe values of the random variable outside a fixed interval about the mean.

Higher order moments usually are not of as much importance to us as are the mean and standard deviation of a process. Generally, however, the odd central moments (i.e., n equal to an odd integer) characterize the skewness or asymmetry

of the random variable probability density function. Density functions which satisfy the symmetry condition

$$f(\mu_y+a) = f(\mu_y-a) \text{ , for any } a \tag{2-18}$$

have identically zero odd central moments. The even central moments, like the second central moment, characterize the dispersion of the distribution. We illustrate the calculation of mean and variance with the following example.

EXAMPLE 2-2

Calculate the mean and variance of a normally distributed random variable with the density expressed in equation (2-11). Explicitly

$$\mu_y = \int_{-\infty}^{\infty} y \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-1/2 \left(\frac{y-\mu}{\sigma}\right)^2\right] dy$$

Observe that the density function is symmetric about μ so that

$$\int_{-\infty}^{\infty} (y-\mu)\, f(y)\, dy = 0$$

and since the integral of the density function is unity

$$\mu \int_{-\infty}^{\infty} f(y)\, dy = \mu = \int_{-\infty}^{\infty} y\, f(y) dy = \mu_y$$

Thus the quantity μ specified in the normal density function is the mean value of the distribution. To obtain the variance we note that:

$$\int_{-\infty}^{\infty} f(y)dy = 1$$

and multiplying both sides by σ we have explicity:

$$\int_{-\infty}^{\infty} 1/\sqrt{2\pi}\ \exp\left[-1/2\left(\frac{y-\mu}{\sigma}\right)^2\right] dy = \sigma$$

Differentiating this result with respect to σ gives

$$\int_{-\infty}^{\infty} \frac{(y-\mu)^2}{\sigma^3\sqrt{2\pi}}\ \exp\left[-1/2\left(\frac{y-\mu}{\sigma}\right)^2\right] dy = 1$$

and multiplying both sides by σ^2 gives

$$\int_{-\infty}^{\infty} (y-\mu)^2\ f(y)\ dy = \sigma^2$$

Thus the quantity σ^2 specified in the normal density function is the variance of the distribution. The normal density function is completely specified in terms of the mean μ_y and the variance σ^2.

Plots of normal density functions having the same mean and different variances are presented in figure 8. Observe that a larger variance of the density function corresponds to a larger dispersion or spread of the density function about the mean.↩

For a given probability experiment, there is no unique random variable having real values assigned to each outcome or event of the experiment. Infinitely many random variables may be defined upon the same experiment. Individually these variables display the properties discussed in section II.3 and II.4 of this report. Together, two or more random variables defined upon the same experiment have additional properties which we shall find useful in our data analysis. Properties of two or more random variables are defined in the next section.

5 PROPERTIES OF SEVERAL RANDOM VARIABLES

Analogous to equation (2-2) we define the joint probability distribution function of two random variables as the probability of the event $\{x \leq y \text{ and } w \leq z\}$ and is written

$$F(y,z) = P(\{x \leq y \text{ and } w \leq z\}) \qquad (2\text{-}19)$$

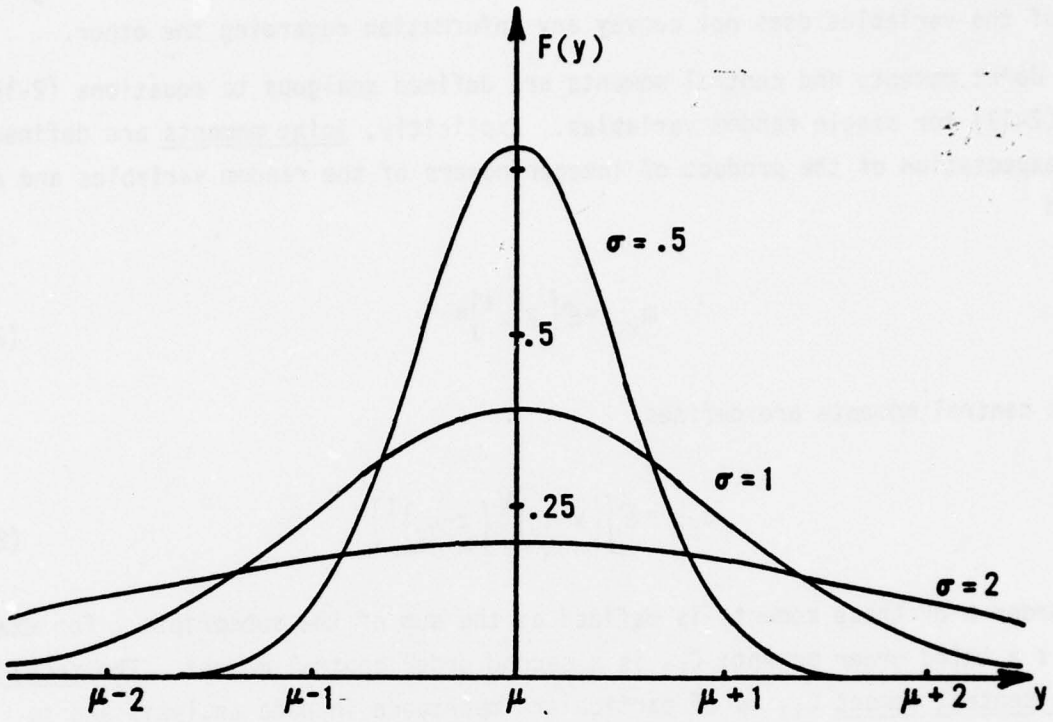


Figure 8. Variance Effects of the Normal Density Function.

Similarly, the joint probability density function is defined as the partial derivative of the joint distribution function with respect to each variable. Notationally,

$$f(y,z) = \frac{\partial^2}{\partial y \partial z} F(y,z) \tag{2-20}$$

The random variables y and z are independent if and only if the joint density function can be expressed in the form*

$$f(y,z) = f_y(y) f_z(z) \tag{2-21}$$

*The subscripts in equation (2-21) distinguish the two density functions $f_y(y)$ for the random variable y and $f_z(z)$ for the random variable z. The subscripts have no other significance.

Independence of two random variables means, practically, that knowledge of one of the variables does not convey any information regarding the other.

Joint moments and central moments are defined analgous to equations (2-16) and (2-17) for single random variables. Explicitly, joint moments are defined as the expectation of the product of integer powers of the random variables and denoted

$$m_{k\ell} = \mathcal{E}\left[y^k z^\ell\right]* \tag{2-22}$$

Joint central moments are defined:

$$C_{k\ell} = \mathcal{E}\left[\left(y-\mu_y\right)^k \left(z-\mu_z\right)^\ell\right] \tag{2-23}$$

The order n of these moments is defined as the sum of the subscripts. For example m_{12} is a third order moment; C_{11} is a second order central moment. The second order central moment C_{11} is of particular importance in data analysis and is called the covariance. For fixed variances of the component random variables, an increase in C_{11} magnitude corresponds to a greater and greater linear dependence between the two random variables. The linear dependence of two random variables is characterized directly by the correlation coefficient ρ_{yz} which is simply C_{11} normalized by the product of the standard deviations of the component processes

$$\rho_{yz} = \frac{C_{11}}{\sigma_y \sigma_z} \tag{2-24}$$

---

*The definition of expectation presented in equation (2-13) must be generalized to the case of two random variables for application to equation (2-22) above. Explicitly the expanded definition is

$$\mathcal{E}[\cdot] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [\cdot]\, f(y,z)\, dy\, dz$$

Expectation means the average with respect to all the component random variables.

Two random variables are linearly dependent if one of the random variables can be expressed as a linear function of the other plus a third independent random variable. In the case that no such linear function exists, the two random variables are linearly independent or uncorrelated. Linearly independent random variables have identically zero correlation coefficients. Thus from equation (2-24) we conclude that:

$$C_{11} = \mathcal{E}\left[\left(y-\mu_y\right)\left(z-\mu_y\right)\right]$$

$$= \mathcal{E}[yz] - \mu_y\mu_z = 0 \tag{2-25}$$

Often equation (2-25) is used to define uncorrelated random variables as those random variables satisfying the condition that

$$\mathcal{E}[yz] = \mu_y\mu_z \tag{2-26}$$

Independent random variables are always uncorrelated since equation (2-26) follows directly from equations (2-21) and (2-22).

We explicitly show the important relationships between correlation coefficients and linear dependence of random variables by the following construction. Assume that random variables y and z with means μ_y and μ_z, respectively, and nonzero variances σ_y^2 and σ_z^2 and a correlation coefficient ρ_{zy} are given. Then we shall show that there exists a random variable w such that

$$w = y-az, \text{ i.e. } y = az + w \tag{2-27}$$

$$a = \frac{\sigma_y}{\sigma_z}\rho \tag{2-28}$$

$$\mu_w = \mu_y - a\mu_z \tag{2-29}$$

$$\sigma_w^2 = \sigma_y^2\left[1 - \rho_{yz}^2\right] \tag{2-30}$$

and

$$\mathcal{E}\left[\left(z-\mu_z\right)\left(w-\mu_w\right)\right] = 0 \tag{2-31}$$

We proceed with the proof of these properties by defining a random variable x and then showing that x satisfies all the properties attributed to w in equations (2-27) through (2-31) above. Because functions of random variables are also random variables, we define x as the difference between y and a scalar multiple b of z.

$$x \triangleq y - bz \tag{2-32}$$

The scalar b is a constant which we conveniently choose so that x satisfies the desired properties. Clearly, by taking expectations of equation (2-32)

$$\mu_x = \mu_y - b\mu_z \tag{2-33}$$

and subtracting this result from equation (2-32), squaring each side, and again taking expectations one obtains:

$$\begin{aligned}\sigma_x^2 &= \mathcal{E}\{[(y-\mu_y) - b(z-\mu_z)]^2\} \\ &= \sigma_y^2 - 2bC_{11} + b^2\sigma_z^2\end{aligned} \tag{2-34}$$

Now we explicitly evaluate the covariance of x and z and equate it to zero by appropriate choice of b.

$$\begin{aligned}\mathcal{E}\left[\left(z-\mu_z\right)\left(x-\mu_x\right)\right] &= \mathcal{E}\left\{\left(z-\mu_z\right)\left[\left(y-\mu_y\right) - b\left(z-\mu_z\right)\right]\right\} \\ &= C_{11} - b\sigma_z^2 \overset{\text{set}}{=} 0\end{aligned} \tag{2-35}$$

from which we obtain:

$$b = C_{11}/\sigma_z^2 = \sigma_y/\sigma_z \left(C_{11}/\sigma_y\sigma_z\right)$$

$$= \frac{\sigma_y}{\sigma_x} \rho_{yz} \tag{2-36}$$

Substitution of this result into equation (2-34) and substitution for C_{11} gives

$$\sigma_x^2 = \sigma_y^2 - \rho_{yz}^2 \sigma_y^2$$

$$= \sigma_y^2 \left[1-\rho_{yz}^2\right] \tag{2-37}$$

Thus making the identification of x with w and b with a and comparing equations (2-32), (2-36), (2-33), (2-37), and (2-35) with equations (2-27) through (2-31), respectively, we have completed the desired proof. These observations follow directly from equations (2-27) through (2-31).

(i) as the correlation between y and z increases, the variance of w, σ_w^2 decreases. For $\rho_{yz} \equiv 1$, $\sigma_w^2 = 0$

(ii) $|\rho_{yz}| \leq 1$ since σ_w^2 must be nonnegative

(iii) $\sigma_w^2 < \sigma_y^2$ and equality holds only if $\rho_{yz}^2 = 0$

(iv) the error in approximating y by a linear function of z is precisely the variance of w; the correlation coefficient squared is precisely the power in y attributable to z normalized by the total power in y. Explicitly,

$$\rho_{yz} = \frac{\mathcal{E}\left[a^2 \left(z-\mu_z\right)^2\right]}{\mathcal{E}\left[\left(y-\mu_y\right)^2\right]} = \frac{a^2\sigma^2_z}{\sigma_y{}^2} \tag{2-38}$$

(v) Zero correlation between random variables y and z implies that the linearity coefficient a of equation (2-27) is identically zero and that $w \equiv y$

(vi) The quantities a, ρ_{yz}, μ_w, and σ_w^2 are uniquely determined by the means, variances, and the covariance of the random variables y and z

The concepts of linear dependence, correlation and joint random variables and their application to a data analysis problem are demonstrated in the next example.

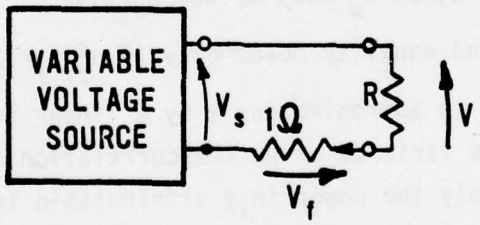
EXAMPLE 2-3

Suppose that we wish to estimate the value of an unknown resistance R by simultaneously measuring the current I through the resistor and the voltage V across it. Then using Ohm's law we obtain the resistance as R = V/I. Further suppose that because of old equipment and coarse meter dials we cannot obtain a sufficiently accurate value for R. We repeat the experiment shown pictorially in Figure 9 for various values of source voltage V_s and make the measurements tabulated below. From these values plotted in figure 10 we wish to extract the "best estimate" of the unknown resistance.

V(volts)	.28	.29	.48	.69	.97	1.07	1.24	1.23	1.53	1.81
I(ma)	.22	.34	.47	.64	.100	1.02	1.18	1.26	1.51	1.90

We formulate a "least-squares" solution by assuming that the measured voltage V is the sum of a linear function of I and a measurement noise term n.

We write: $\hat{V} = \hat{R}I + b$



$I = V_f$

Figure 9. Experimental Resistance Measurement for Example 2-3.

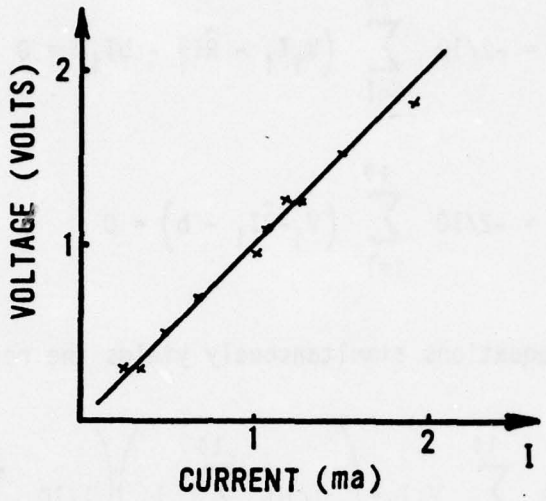


Figure 10. Mean Square Estimation of a Voltage Current Relationship.

where the caret above V and R denotes that they are estimated quantities. The constant b denotes instrument bias. Since we have assumed $\hat{V}$ differs from V by a noise term, we have

$$n_i = V_i - \hat{V}_i$$

We arbitrarily choose the unknown parameters $\hat{R}$ and b so as to minimize the noise variance, or equivalently, we minimize the mean-square error, by which an estimated quantity $\hat{V}$ differs from the actual measurement V.

We define a performance measure

$$E = \frac{1}{10} \sum_{i=1}^{10} \left[V_i - \left(\hat{R} I_i + b \right) \right]^2$$

and we minimize E by optimum choice of the unknown parameters $\hat{R}$ and b. The solution of the unconstrained minimization problem is most readily obtained by equating the partial derivatives of E with respect to $\hat{R}$ and of E with respect to b both equal to zero and solving these equations simultaneously for the unknowns R and b. Executing these steps

$$\partial E/\partial \hat{R} = -2/10 \sum_{i=1}^{10} \left(V_i I_i - \hat{R} I_i^2 - b I_i\right) = 0$$

$$\partial E/\partial b = -2/10 \sum_{i=1}^{10} \left(V_i - \hat{R} I_i - b\right) = 0$$

and solving the above equations simultaneously yields the result

$$\hat{R} = \frac{1/10 \sum_{i=1}^{10} V_i I_i - \left(1/10 \sum_{i=1}^{10} I_i\right)\left(1/10 \sum_{i=1}^{10} V_i\right)}{D} = 9.6\text{K}\Omega$$

$$b = \frac{\left(-1/10 \sum_{i=1}^{10} I_i\right)\left(1/10 \sum_{i=1}^{10} V_i I_i\right) + \left(1/10 \sum_{i=1}^{10} I_i^2\right)\left(1/10 \sum_{i=1}^{10} V_i\right)}{D}$$

$$= 0.04 \text{ volts}$$

where $$D = 1/10 \sum_{i=1}^{10} I_i^2 - \left(1/10 \sum_{i=1}^{10} I_i\right)^2$$

In view of the relative frequency interpretation of probabilities developed in section II.2, we may re-interpret the above equations in terms of the expectations they approximate. We have

$$1/10 \sum_{i=1}^{10} I_i \approx \mu_I$$

$$1/10 \sum_{i=1}^{10} V_i \approx \mu_V$$

$$1/10 \sum_{i=1}^{10} I_i^2 \approx \mathcal{E}[I^2] = \sigma_I^2 + \mu_I^2$$

$$1/10 \sum_{i=1}^{10} V_i I_i \approx \mathcal{E}[VI] = C_{11} + \mu_V \mu_I$$

Thus expressing $\hat{R}$ and b in terms of these approximations gives

$$\hat{R} \approx C_{11}/\sigma_I^2 = \frac{\sigma_V}{\sigma_I} \rho_{VI}$$

$$b \approx \frac{\mu_V \sigma_I^2 - \mu_I C_{11}}{\sigma_I^2} = \mu_V - \hat{R}\, \mu_I$$

Had we known the statistics of V and I we could have simply applied equation (2-28) to have calculated R in terms of σ_V, σ_I, and ρ_{VI}. Typically, however, these quantities are not known à priori but must be approximated. Equivalently, the mean-square-error procedure developed in this example yields the best answer given the available data.←

Thus far the concepts of randomness and probability, random variables and their characterization by probability density or distribution functions have been developed. The mean or average value and the variance or dispersion of the random variable have been defined and shown to be two of the more important characterizations. Covariance and the correlation coefficient are shown to be important properties of random variables whenever functions of random variables are of interest.

6 STOCHASTIC PROCESSES

The concept of random variables introduced in section II.3 has no time dependence associated with it. This limitation is serious when one is most interested in the effect of random phenomena upon systems having input-output relationships described by differential equations. In this section, the concept of stochastic processes is introduced to allow analysis of those situations in which time dependent random effects excite linear dynamical systems or corrupt a time dependent information signal we wish to examine.

A time function $x(t)$ having numerical values which depend not only upon the value of its argument t but also upon the outcome of probability experiment is called a stochastic process. For example the function

$$x(t) = \begin{cases} \sin wt \ , & \text{if HEAD's} \\ \cos wt \ , & \text{if TAIL's} \end{cases} \tag{2-39}$$

where HEADS or TAILS is the outcome of a coin toss experiment as in section II.2 is an example of one stochastic process. In practical cases of interest, the underlying probability space characterizes the size distribution of atmospheric turbules or the "shot" rate of electronic noise sources or some other physical phenonema having outcomes which cannot be precisely predicted in advance but which do having varying effects upon our particular experiment. Often, the underlying probability space is not well characterized, known, or understood other than through observation of the resulting signal or effect on our experiment. The observed effect is also a stochastic process because its value depends upon a stochastic process excitation which in turn depends upon the outcome of the underlying probability experiment outcome. The observed stochastic process may be, for example, the intensity fluctuations I(t) of an electromagnetic wave propagated through a turbulent atmosphere or the voltage variations of an electronic system excited by noise.

Random signal analysis techniques, which are the primary topic of this guide, are in essence a collection of methods for characterizing these stochastic processes and the transform relationships which model the interaction of stochastic signals with linear dynamical systems.

Basically, stochastic processes are defined on probability experiments in the same manner as random variables. In fact, for fixed agrument t stochastic

processes degenerate to random variables. Characterizations of random variables are adequate to model certain fixed time properties of stochastic processes but new concepts, definitions and characterizations are required to parameterize important additional properties of the stochastic process.

Recall that a stochastic process defines a real function of the independent variable t for each outcome of a probability experiment. The collection of all possible time functions is the ensemble of the stochastic process. Both the member of the ensemble (or equivalently the outcome of the experiment) and time must be known to assign a value to the stochastic process function. A stochastic process ensemble is partially depicted in figure 11 which shows

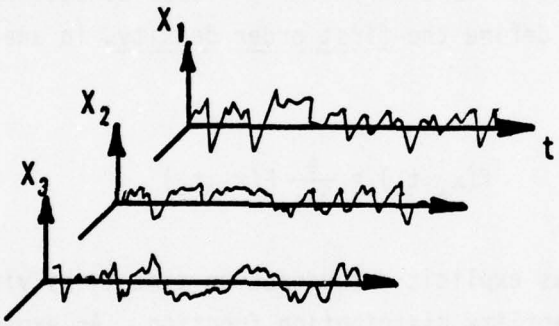


Figure 11. A Stochastic Process Ensemble.

explicitly the dual dependence of the stochastic process. The properties mean, variance, n^{th} order moments and central moments, joint moments, covariance, and correlation developed in sections II.4 and II.5 for random variables are equally applicable to a stochastic process at each fixed time. Consequently, these properties are defined as functions of time for stochastic processes. Additional properties of the stochastic process to be developed in this section are generalizations of the basic properties that have already been introduced. Stochastic processes are completely characterized by joint probability distribution functions of the form:

$$F\left(x_1, x_2, \ldots\ldots, x_n, t_1, t_2, \ldots t_n\right) =$$

$$P\left(\{x(t_1) \leq x_1, x(t_2) \leq x_2, \ldots x\left(t_n\right) \leq x_n\}\right) \qquad (2\text{-}40)$$

which must be known for all finite sets of $\{t_i\}$, (reference 3, p. 51). Such functions are cumbersome or defy specification in all but the simplest cases. Fortunately, several other properties of stochastic processes -- namely stationarity, ergodicity, and normality -- are generally valid and allow a greatly simplified characterization. These properties are precisely defined subsequent to the following preliminary definitions.

7. MEANS, MOMENTS, AND ERGODICITY

The concepts of means and moments developed in sections II.4 and II.5 for random variables are explicitly extended in this section to stochastic processes. First, probability density functions and expectation operators are defined for stochastic processes. The probability distribution function equation (2-40) for stochastic processes is considered for the case consisting of precisely a single element t_1. We define the first order density, in analogy with equation (2-3), as

$$f(x_1,t_1) = \frac{\partial}{\partial x_1} F(x_1,t_1) \tag{2-41}$$

The density $f(x_1,t_1)$ has explicit dependence on time t_1 by virtue of the time dependence of the probability distribution function. An expectation operator follows from equation (2-41) as a straightforward generalization of expectation for a single random variable as defined in equation (2-13).

$$\mathcal{E}[\cdot] = \int_{-\infty}^{\infty} [\cdot]\, f(x,t)\, dx \tag{2-42}$$

The mean is defined as the expectation of the process itself.

$$\mu_x(t) = \mathcal{E}[x(t)] = \int_{-\infty}^{\infty} x(t) f(x,t) dx \tag{2-43}$$

The mean at any fixed time is simply the average of x over the elements of the ensemble.

Second-order moments are defined in terms of the second-order density function

$$f(x_1, x_2, t_1, t_2) = \frac{\partial}{\partial x_1 \partial x_2} F(x_1, x_2, t_1, t_2) \tag{2-43a}$$

The correlation function is defined as the second moment

$$m_x(t_1,t_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1(t_1)x_2(t_2)\, f(x_1, x_2, t_1, t_2)\, dx_1\, dx_2$$

$$= \mathcal{E}[x(t_1)\, x(t_2)] \tag{2-44}$$

and the covariance is defined as

$$C_x(t_1,t_2) = \mathcal{E}\{[x(t_1) - \mu_x(t_1)]\,[x(t_2)-\mu_x(t_2)]\}$$

$$= m_x(t_1,t_2)\, \mu_x(t_1)\, \mu_x(t_2) \tag{2-45}$$

The functions of m_x and C_x are defined in terms of a single stochastic process x and therefore they are often called autocorrelation and autocovariance functions, respectively. Correlation and covariance functions between two stochastic processes x and y are defined as

$$m_{xy}(t_1,t_2) = \mathcal{E}[x(t_1)\, y(t_2)] \tag{2-46}$$

and

$$C_{xy}(t_1,t_2) = \mathcal{E}\{[x(t_1) - \mu_x(t_1)][y(t_2) - \mu_y(t_2)]\}$$

$$= m_{xy}(t_1,t_2) - \mu_x(t_1)\mu_y(t_2) \tag{2-47}$$

Stationarity defines the property that the probabilities describing our process do not depend upon the time origin or the time frame in which we describe our process. For example, stationarity implies that the probability density function satisfies the equalities indicated in the following cases,

$$f(x_1,t_1) \equiv f(x_1,t_1+\tau)$$

$$f(x_1,x_2,t_1,t_2) \equiv f(x_1,x_2,t_1+\tau,t_2+\tau) \tag{2-48}$$

which must be valid for any and all τ. Clearly, then $f(x_1,t_1)$ is independent of t_1 and may be more simply written $f(x_1)$. Similarly, $f(x_1, x_2, t_1, t_2)$ is independent of the specific time reference used and therefore depends only upon the relative time difference $\tau = t_1 - t_2$. Thus for stationary processes,

$$f(x_1, x_2, t_1, t_2) \equiv f(x_1, x_2, \tau) \tag{2-49}$$

Stationarity reduces the complexity of the probabilistic characterization. Stationary processes have means μ_x which are time-invariant. The correlation and covariance functions, like the stationary second-order density functions, depend only upon the time difference τ. We denote these functions $m_{xy}(\tau)$ and $C_{xy}(\tau)$ for the stationary process correlation and covariance functions, respectively. We shall consider only stationary processes throughout the remainder of the guide.

Expectations have been defined as averages over the underlying probability experiment. In signal analysis, another important average is time-average. Time-average mean, correlation and covariance quantities are defined for a particular realization x_i and y_i of the stochastic processes as

$$\overline{x_i} = \operatorname*{\ell im}_{T\to\infty} \; 1/T \int_{-T/2}^{T/2} x_i(t)dt \tag{2-50}$$

$$R_{xy}(\tau) = \underset{T\to\infty}{\ell im}\ 1/T \int_{-T/2}^{T/2} x_i(t+\tau)y_i(t)dt \qquad (2\text{-}51)^*$$

$$\bar{C}_{xy}(\tau) = \underset{T\to\infty}{\ell im}\ 1/T \int_{-T/2}^{T/2} \left[x_i\ (t+\tau) - \bar{x}_i\right]\left[y_i(t)-\bar{y}_i\right]dt \qquad (2\text{-}52)$$

provided that the indicated limits exist. The time-averages defined above are random variables since they have values dependent upon the particular realization x_i and y_i of the underlying stochastic process (reference 4, p. 326). Assuming process stationarity and taking expectations of the above quantities we find that

$$\mathcal{E}\left[\bar{x}_i\right] = \underset{T\to\infty}{\ell im}\ 1/T \int_{-T/2}^{T/2} \left[x_i(t)\right]dt$$

$$= \mu_x \qquad (2\text{-}53)$$

Similarly,

$$\mathcal{E}\left[R_{xy}(\tau)\right] = m_{xy}(\tau) \qquad (2\text{-}54)$$

$$\mathcal{E}\left[\bar{C}_{xy}(\tau)\right] = C_{xy}(\tau) \qquad (2\text{-}55)$$

---

*In the literature $R_{xy}(\tau)$ is also defined as

$$R_{xy}(\tau) = \underset{T\to\infty}{\ell im}\ 1/T \int_{-T/2}^{T/2} x_i(t)\ y_i(t+\tau)\ dt.$$

The definition is simply a matter of convention and does not substantively change the results. The reader is cautioned, however, to familiarize himself with the particular convention adopted by another author.

For many stochastic processes, the random variables $\bar{x}_i$, $R_{xy}(t)$ and $\bar{C}_{xy}(t)$ have zero variance, have identical values independent of the stochastic process realization x_i and y_i used in their time-average calculation, and are identically equal to the corresponding ensemble average. Should these three properties hold for all average values of a stochastic process, then the process is said to be ergodic. An important consequence of ergodicity is that each element of the stochastic process ensemble is representative of the ensemble as a whole. Ergodicity allows us to compute important characterizations (mean, covariance, etc.) of a stochastic process by applying time-average calculations to a simple observation of the ensemble. We shall consider only ergodic stochastic processes throughout the remainder of this guide. We state without proof that most of the stochastic processes encountered in engineering practice are ergodic or nearly ergodic in the sense that our analysis techniques remain valid and useful.

8 LINEAR DYNAMIC SYSTEMS

A linear dynamic system is a system in which certain dependent (output) variables satisfy a given linear differential equation having one or more independent variables or forcing functions. We shall consider the general case in which these forcing functions are stochastic processes and derive stochastic process characterizations of the system output in terms of the input and the system transfer function or differential equation. Having completed this characterization we show the importance or correlation and covariance calculations in estimating dynamic system input-output relationships. We assume the reader has a knowledge of differential equations and Fourier transforms such as presented by DeRusso (reference 6, Chapters 1 through 4). We briefly introduce these topics to define notation.

It can be shown (reference 6, p. 23) that a time-invariant linear differential equation has an output of y(t) expressed as a function of the input x(t) by the equation

$$y(t) = \int_{-\infty}^{\infty} h(t-\tau)\, x(\tau)\, dt = \int_{-\infty}^{\infty} h(\tau)\, x(t-\tau)\, dt \qquad (2\text{-}56)$$

where $h(\tau)$ is the impulse response of the system. Such a system is depicted diagrammatically in figure 12.

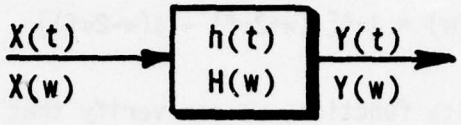


Figure 12. A Linear Dynamic System.

The Fourier transform X(w) of a signal x(t) is a complex function representing the magnitude and phase of the frequency components of x(t). The transform is defined as

$$F[x(t)] \triangleq X(w) = \int_{-\infty}^{\infty} x(t)\exp(-jwt)dt^* \tag{2-57}$$

An inverse transformation also exists by which x(t) may be obtained from X(w).

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(w)\ \exp(jwt)dw \tag{2-58}$$

Observe that the transform is a linear operator.

EXAMPLE 2-4

We explicitly calculate the Fourier transform of $x(t) = A \sin(2\pi ft)$ to show that the transform represents the frequency or spectral content of the corresponding time signal. Strictly speaking, the Fourier transform of this particular x(t) does not exist since x(t) is not absolutely integrable, i.e.,

$$\int_{-\infty}^{\infty} |A\sin(2\pi ft)|\ dt = \infty$$

*The quantity j is the square root of -1, $j = \sqrt{-1}$.

Consequently we cannot apply equation (2-57) directly to evaluate X(w). Observe, however, that assuming

$$X(w) = j\pi A[\delta(w+2\pi f) - \delta(w-2\pi f)]$$

where δ is the Dirace delta function, we can verify that X(w) is indeed the transform of x(t). We evaluate the inverse transform by use of the delta function sifting property (reference 6) to obtain:

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} X(w)\exp(jwt)\,dw = \frac{A}{2} j[\exp(-j2\pi ft) - \exp(j2\pi ft)]$$

$$= A \sin(2\pi ft) = x(t)$$

This transform pair is shown in figure 13. Observe that as the frequency of the sinusoid increases, the transform delta functions move farther from the origin.

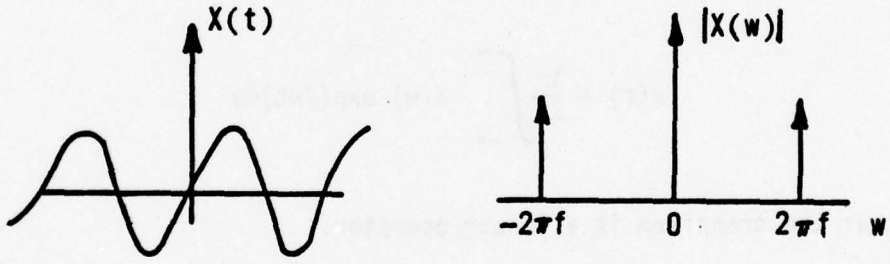


Figure 13. Fourier Transform Pairs.

The linear dynamical system input-output relationship given in equation (2-56) can be Fourier transformed to yield an equivalent transform description of the system. Performing the transform, we have

$$Y(w) = \int_{-\infty}^{\infty} y(t)\exp(-jwt)\,dt$$

$$= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} h(t-\tau)\,x(\tau)\,d\tau\right]\exp(-jwt)\,dt$$

We interchange the order of integration and multiply the integrand by the quantity

$$\exp(jw\tau)\exp(-jw\tau) = 1$$

and obtain

$$Y(w) = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} h(t-\tau)\exp[-jw(t-\tau)]\,dt\right]\exp(-jw\tau)\,x(\tau)\,d\tau$$

$$= H(w)\,X(w) \qquad (2\text{-}59)$$

Equation (2-59) shows the simple relationship that the Fourier transform of the output of a linear system is the product of the Fourier transforms of the input signal and the system impulse response function. A similar input-output relationship is now shown for the system driven by stochastic forcing functions. Since Fourier transform techniques are <u>not</u> applicable to stochastic process (they are not absolutely integrable), a few preliminary concepts must be defined prior to the formulation of a stochastic process transfer function.

Signal frequency distributions for stochastic processes are derived by the following heuristic argument. Given any stochastic process x(t) we may filter that process with a narrow band filter having center frequency f_0 and bandwidth Δf. The transform of an ideal bandpass filter is plotted in figure 14. The

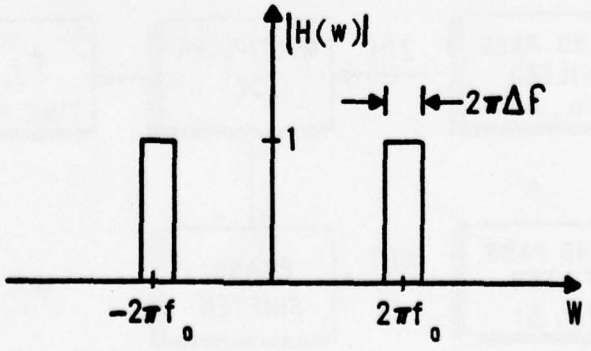


Figure 14. An Ideal Bandpass Filter Transform Characteristic.

output signal of this bandpass filter, just as in the analogous case in which a deterministic signal is filtered, represents the input signal spectral content within the passband frequencies. The power of the filter output represents that portion of the input signal power within the passband of the filter. We define signal <u>power</u> P as the time average signal squared for both deterministic and stochastic signals. This time-average for stochastic signals is the autocorrelation function evaluation for zero argument

$$P = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x^2(t)\, dt$$

$$= m_{xx}(0) \quad \text{(for stochastic processes)} \qquad (2\text{-}60)$$

The power output of the filter can also be measured by time-averaging the signal squared. The complete analysis procedure of bandpass filtering, squaring, and time-averaging is depicted in figure 15a. We denote the average <u>power density</u>

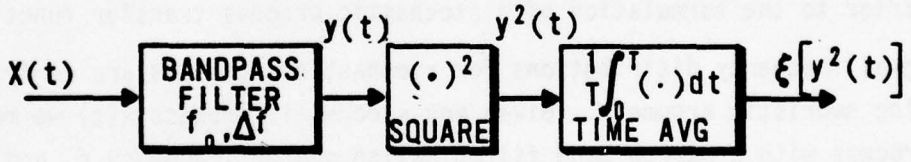


Figure 15a. Measuring Narrow Band Signal Power.

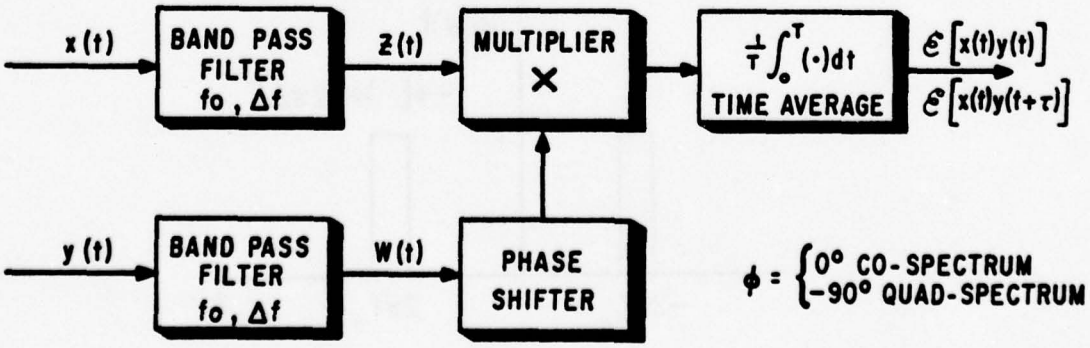


Figure 15b. Measuring Narrow Band Cross-Spectral Power.

of the stochastic process as $P(x,f_o,f)$ and define it as the power output of the bandpass filter divided by the filter bandwidth (see figure 15).

$$\Phi\ (x,f_o,\Delta f) = R_{yy}(o)/\Delta f \tag{2-61}$$

Conceptually these ideas are extended to the use of infinitely many contiguous bandpass filters having infinitesimally small bandwidths Δf. The power spectral density (PSD), $\Phi_{xx}(f)$, of a stochastic process is defined as the limit as Δf approaches zero of the average power density and we write:

$$\Phi_{xx}(f) = \lim_{\Delta f \to o} P(x,f,\Delta f) \tag{2-62}$$

Cross-Spectral Densities (CSD), $\Phi_{xy}(f)$, between two signals $x(t)$ and $y(t)$ are analogously defined but must be generalized in order to characterize the important phase relationships between two distinct signals that do not arise in considerations of a single signal. For PSD measurements, a signal $x(t)$ is passed through a narrow-band bandpass filter which gives a sinusoidal output signal $y(t)$ having an amplitude proportional to the square-root of the signal power within the filter pass-band. The power is determined by the time-averaged filter output squared. The cross-spectral power between signals $x(t)$ and $y(t)$ is calculated by multiplying the outputs of two narrow band-filters excited by $x(t)$ and $y(t)$, respectively; then the time average product is computed as shown in figure 15b. Note that this procedure gives the proper PSD in the case that $y(t)$ equals $x(t)$. For more general signals of $x(t)$ and $y(t)$, it represents the in-phase signal power common to signals $x(t)$ and $y(t)$. An out-of-phase component, the quadspectrum, must also be computed to complete the two signal common characterization. The requirement for calculation of the quadspectrum follows from consideration of the case in which $y(t)$ is a delayed or phase-shifted version of $x(t)$ such that the phase shift $y(t)$ with respect to $x(t)$ is precisely 90°. In this case the averaged input is the product of orthogonal signals sine and cosine and the output will be zero. By shifting one of the input signals 90° at the pass band center frequency, f_o, the roles of the in-phase cospectrum and the out-of-phase quadspectrum are reversed. Thus by inserting a 90° phase shifter in the $y(t)$ signal path, the quadspectrum between x and y is determined by procedures otherwise identical to those employed to determine the cospectrum.

Denoting the cospectrum power by P_c we have that

$$P_c\left(x,y,f_o,\Delta f\right) = R_{zw}(o)/\Delta f \tag{2-63a}$$

and analogously the quadspectrum power P_q is

$$P_q\left(x,y,f_o,\Delta f\right) = R_{zw}(\tau)/\Delta f \tag{2-63b}$$

where τ represents the time shift required to change the relative phase of x with respect to y exactly 90° at the passband center frequency.

$$2\pi f_o \tau = \pi/2$$

or

$$\tau = \frac{1}{4f_o}$$

$R_{zw}(t)$ represents the cross correlation function of signals z and w obtained by bandpass filtering signals x and y respectively. The cross-spectral density $\Phi_{xy}(f)$ is defined in terms of the co- and quad-spectrum as

$$\Phi_{xy}(f) = \lim_{\Delta f \to o} \left[P_c\left(x,y,f_o,\Delta f\right) - jP_q\left(x,y,f_o,\Delta f\right)\right] \tag{2-64}$$

Unlike the power calculations for PSDs which must always yield a non-negative result, P_c and P_q may assume any real value. Also note that, in general, the cross-spectral density is a complex function of frequency. These generalizations over the PSD are required to account for the phase relationships betwen two signals.

PSDs and CDS are shown to be directly related to Fourier transforms of the auto and cross-covariance functions. By direct calculation

$$F(w) = \int_{-\infty}^{\infty} R_{xy}(\tau) \exp(-jw\tau)\, d\tau$$

$$= \lim_{T\to\infty} \frac{1}{T} \int_{T/2}^{\infty} \int_{-\infty}^{T/2} x(t+\tau)\, y(t)\, dt \exp(-jw\tau)\, d\tau \qquad (2\text{-}65)$$

For mathematical convenience we define a truncated function $x_T(t)$

$$x_T(t) = \begin{cases} x(t) & |t| < T/2 \\ 0 & \text{otherwise} \end{cases} \qquad (2\text{-}66)$$

and defining $y_T(t)$ similarly, substitute these relations into equation (2-65) (and ignoring negligible end effects we have)

$$F(w) = \lim_{T\to\infty} 1/T \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_T(t+\tau)\, y_T(t) dt \exp(-jw\tau) d\tau \qquad (2\text{-}67)$$

Multiplying the integrand by exp (-jwt) exp (+jw) = 1 and rearranging gives

$$F(W) = \lim_{T\to\infty} 1/T \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} x_T(t+\tau)\exp[-jw(t+\tau)]dt \right\} \underset{y_T(t)}{} \exp(jw\tau) d\tau$$

$$= \lim_{T\to\infty} \frac{X_T(w)\, Y_T(-w)}{T} \qquad (2\text{-}68)$$

Recall from equation (2-58) and Example 2-4 that the Fourier transform represents a signals spectral content. From Parseval's identity for Fourier transform pairs, namely,

$$\int_{-\infty}^{\infty} x(t)\, y(t)dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(w)\, Y(-w)dw \tag{2-69}$$

we see that the Fourier transform magnitude squared represents precisely the signal's energy distribution in frequency. Specifically we have

$$\int_{-\infty}^{\infty} x^2(t)dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(w)\, X(-w)dw \tag{2-70}$$

but

$$X(w)X(-w) = |X(w)|^2$$

Consequently,

$$\frac{1}{2\pi}\left[\int_{-2\pi\,(f_o+\Delta f/2)}^{-2\pi\,(f_o-\Delta f/2)} |X(w)|^2dw + \int_{2\,(f_o-\Delta f/2)}^{2\pi(f_o+\Delta f/2)} |X(w)|^2dw\right]$$

represents a signal energy in a f frequency band about center frequency f_o. These concepts for signal energy properties are also applicable to signal power considerations because energy normalized by time is power. Consequently,*

*A subtlety in the concept of frequency distribution of power has arisen upon which we briefly elaborate here. The bandpass conceptualization of power in equation (2-61) and equation (2-62) is valid for positive frequencies only. We cannot build a filter having only negative bandpass frequencies. Any realizable filter passes equally frequencies $f_2 = -f_1$. Thus we define PSD and CSD for positive frequencies only. However, Fourier transforms define spectral content for both positive and negative frequencies, hence

$$\frac{1}{2\pi}\int_{-\infty}^{0} \frac{|X_T(w)|^2}{T}\,dw = \frac{1}{2\pi}\int_{0}^{\infty} \frac{|X_T(w)|^2}{T}\,dw = \frac{1}{2}\int_{0}^{\infty} \Phi_{xx}(f)df$$

and equations (2-71) and (2-72) follow.

$$\Phi_{xy}(w/2\pi) = 2 \lim_{T\to\infty} \frac{X_T(w)\ Y_T(-w)}{T}$$

$$= 2\,F\,[R_{xy}(\tau)] \triangleq 2\,G_{xy}(w) \qquad w > 0 \qquad (2\text{-}71)$$

as a special case

$$\Phi_{xx}(w/2\pi) = 2 \lim_{T\to\infty} \frac{|X_T(w)|^2}{T}$$

$$= 2\,F\,[R_{xx}(\tau)] \triangleq 2\,G_{xx}(w) \qquad w > 0 \qquad (2\text{-}72)$$

The characterization tools that have now been developed allow us to calculate input-output signal relationships for linear dynamic systems. Equation (2-56) repeated below describes a linear system output signal y(t) in terms of the system impulse response h and the system input x. This equation is equally valid for stochastic process excitation of the system.

$$y(t) = \int_{-\infty}^{\infty} h(\tau_1)\ x(t-\tau_1)\ d\tau_1 \qquad (2\text{-}73)$$

Taking expectations of equation (2-73) gives the system output mean as a function of the input mean. For stationary processes the output mean is simply a scalar multiple of the input mean.

$$\mu_y = \mathcal{E}[y(t)] = \int_{-\infty}^{\infty} h(\tau_1)\mu_x\ d\tau_1$$

$$= \mu_x \int_{-\infty}^{\infty} h(\tau_1)\ d\tau_1$$

$$= \mu_x\, H(w) \Big|_{w=0}$$

Hence, the average value of the system output is simply the average value of the system input multiplied by the dc gain H(o) of the linear system.

Higher order moments of the system output and joint moments of the system output and input are similarly obtained as functions of the linear system impulse response function h(t) and moments of the input signal. For example, the cross-correlation of the system input and output signals is obtained as follows. Multiplying both sides of the equation (2-73) by $x(t + \tau)$ and taking expectations yields:

$$R_{xy}(\tau) = \int_{-\infty}^{\infty} h(\tau_1)\, R_{xx}(\tau+\tau_1)\, d\tau_1 \tag{2-74}$$

This equation shows that the cross-correlation of the system input and output signals is given as the convolution of the system input signal auto-correlation function with the system impulse response. Taking the Fourier transform of both sides of equation (2-74) gives:

$$\Phi_{xy}(w/2\pi) = H(-w)\, \Phi_{xx}(w/2\pi) \tag{2-75}$$

Similarly, multiplying both sides of equation (2-73) by $y(t + \tau)$ and taking exceptions gives:

$$R_{yy}(\tau) = \int_{-\infty}^{\infty} h(\tau_1)\, R_{yx}(\tau+\tau_1)d\tau_1 \tag{2-76}$$

and transforming:

$$\begin{aligned}\Phi_{yy}(w/2\pi) &= H(-w)\, \Phi_{yx}(w/2\pi) \\ &= H(w)\, \Phi_{xy}(w/2\pi)\end{aligned} \tag{2-77}$$

By the definition of correlation functions and their transforms we have the following properties:

$$R_{xy}(\tau) = R_{yx}(-\tau)$$

$$\Phi^*_{xy}(w/2\pi) = \Phi_{yx}(w/2\pi) = \Phi_{xy}(-w/2\pi) \tag{2-78}$$

Thus substituting for Φ_{yx} in equation (2-77) from equation (2-75) by using equation (2-78) gives the desired result.

$$\Phi_{yy}(w/2\pi) = H(-w)\ \Phi^*_{xy}(w/2\pi)$$

$$= H(-w)[H(-w)\phi_{xx}(w/2\pi)]^*$$

$$= |H(w)|^2\ \Phi_{xx}(w/2\pi) \tag{2-79}$$

since Φ_{xx} is real.

We also write

$$\Phi_{yy}(f) = |H(2\pi f)|^2\ \Phi_{xx}(f) \tag{2-80}$$

This equation shows that a linear dynamical system output signal PSD is precisely the input PSD 'spectrally shaped' by the frequency domain transfer function of the system. We illustrate the response of a linear dynamical system excited by a stochastic process with an example.

EXAMPLE 2-5

A simplified linear dynamic system is considered. The system is modeled by the first order linear time-invariant differential equation

$$\frac{dy}{dt} = ay + ax$$

The quanity y represents the system output while x represents a stochastic process excitation of the system. The differential equation specified above represents a lowpass filter. Spectral components of x having frequencies f less than the filter cutoff frequency $f_c = a/2\pi$ are passed unattenuated to the output y. Spectral components of x at frequencies greater than f_c are attenuated. The linear system impulse response is

$$h(t) = a \exp(-at) \quad t \geq o$$

and its Fourier transform, also known as the system transfer function, is

$$H(w) = (1 + jw/a)^{-1}$$

We specify that the stochastic excitation is a zero mean process with an autocorrelation function

$$R_{xx}(\tau) = q\delta(\tau)$$

This stochastic process is of particular importance in analysis. Since its autocorrelation function is zero for all τ not equal to zero, we conclude that knowledge of the value of x at time t is uncorrelated with its value at any other time. Hence we cannot predict future values of x(t) from past observations. Stochastic processes having autocorrelation functions of the form $q\delta(t)$ are called white noise. This terminology follows from its spectral density, which is constant for all frequencies, and the analogy with white light, which is composed of light of all colors or frequencies. The excitation spectral density is derived from equation (2-72) and $R_{xx}(\tau)$ as

$$\Phi_{xx}(w/2\pi) = 2q \; ; \; w \geq o$$

Using equations (2-75) and (2-80) we calculate that

$$\Phi_{xy}(w/2\pi) = H(-w) \; \Phi_{xx}(w/2\pi)$$

$$= 2q \; (1 - jw/a)^{-1}$$

and

$$\Phi_{yy}(w/2\pi) = |H(w)|^2 \; \Phi_{xx}(w/2\pi)$$

$$= 2q \; [1 + (w/a)^2]^{-1}$$

the correlation functions $R_{xy}(\tau)$ and $R_{yy}(\tau)$ may be calculated in two ways:

(1) their Fourier transforms may be deduced from Φ_{xy} and Φ_{yy} and inverted or

(2) they can be calculated directly from h(t) and $R_{xx}(\tau)$ by use of equations (2-74) and (2-76). We shall calculate $R_{xy}(\tau)$ and $R_{yy}(\tau)$ by both methods to illustrate methodology and to demonstrate the equivalence of the alternative approaches. Since

$$\Phi_{xy}(w/2\pi) = 2q[1-jw/a]^{-1}$$

we conclude that

$$F[R_{xy}(\tau)] = q[1-jw/a]^{-1}$$

which we inverse transform (consult reference 6 or any convenient table of Fourier transform pairs) to obtain

$$R_{xy}(\tau) = aq \exp(a\tau) \quad \tau \leq 0$$

Observe that $R_{xy}(\tau) = 0$ for $\tau > 0$. This follows directly from the causalty of the system output which can be correlated only with <u>past</u> values of the input.

Similarly

$$F[R_{yy}(\tau)] = q[1+(w/a)^2]^{-1}$$

$$= q/2[(1+jw/a)^{-1} + (1-jw/a)^{-1}]$$

and

$$R_{yy}(\tau) = q\frac{a}{2} \exp(-a|\tau|)$$

By the alternative method we calculate

$$R_{xy}(\tau) = \int_{-\infty}^{\infty} h(\tau_1)\, R_{xx}(\tau+\tau_1)d\tau_1$$

$$= q\, \hat{h}(-\tau)$$

$$= aq \exp(a\tau) \qquad \tau \leq 0$$

and

$$R_{yy}(\tau) = \int_{-\infty}^{\infty} h(\tau_1) R_{yx}(\tau+\tau_1)d\tau_1$$

$$= \int_{\max(0,-\tau)}^{\infty} qa^2 \exp(a\tau_1) \exp[-a(\tau+\tau_1)]d\tau_1$$

$$= qa^2 \exp(-a\tau) \int_{\max(0,-\tau)}^{\infty} \exp(-2a\tau_1)d\tau_1$$

$$= q\frac{a}{2} \exp(-a|\tau|)$$

The same results have been obtained by each method available for evaluating the crosscorrelation functions. The input and output signal auto-correlation functions $R_{xx}(\tau)$ and $R_{yy}(\tau)$ are plotted in figure 16.

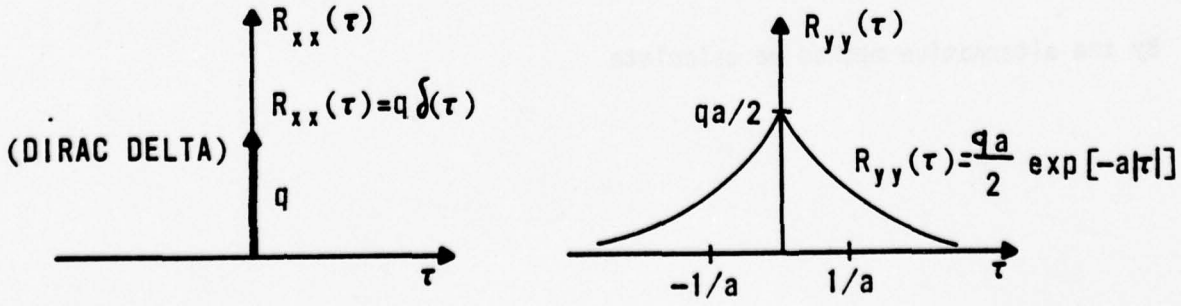


Figure 16. Linear System Input and Output Autocorrelation Functions.

The corresponding power spectral densities are plotted in figure 17. The parameter 1/a represents the correlation time of the output stochastic process and

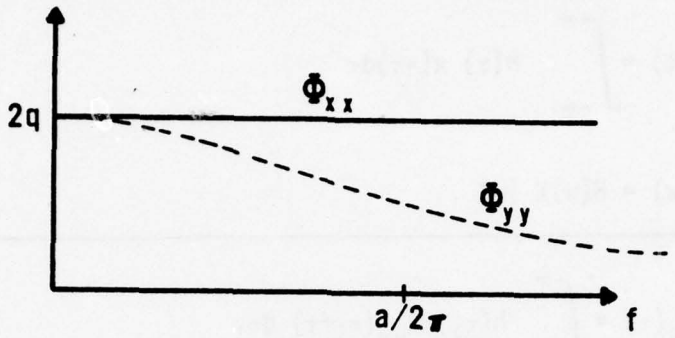


Figure 17. Linear System Input and Output Power Spectral Densities.

is identical to the time constant of the linear system when excited by white noise.

The important relationships developed in this section are summarized in table 1. Also included in the table are other important equalities which the reader is invited to validate. These equations play a fundamental role throughout the remainder of this guide.

Table 1

CORRELATION FUNCTIONS AND SPECTRAL DENSITY RELATIONSHIPS

Definitions		
	$R_{xx}(\tau) = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} x(t+\tau)\, x(t)\, dt$	Autocorrelation
	$R_{xy}(\tau) = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} x(t+\tau)\, y(t)\, dt$	Crosscorrelation
	$\phi_{xx}(w/2\pi) = 2\, F[R_{xx}(\tau)]$ $\quad w \geq 0$	PSD
	$\phi_{xy}(w/2\pi) = 2\, F[R_{xy}(\tau)]$ $\quad w \geq 0$	CSD

Table 1 (Continued)

LINEAR SYSTEM RELATIONSHIPS

System Definitions	$y(t) = \int_{-\infty}^{\infty} h(\tau)\, x(-\tau)d\tau$
	$Y(w) = H(w)X\,(w)$
Correlation Function Relationships	$R_{xy}(\tau) = \int_{-\infty}^{\infty} h(\tau_1)\, R_{xx}(\tau_1+\tau)\, d\tau_1$
	$R_{yx}(\tau) = \int_{-\infty}^{\infty} h(\tau_1)R_{xx}(\tau_1-\tau)d\tau_1$
	$= R_{xy}(-\tau)$
	$R_{yy}(\tau) = \int_{-\infty}^{\infty} h(\tau_1)R_{yx}(\tau_1+\tau)d\tau_1$
	$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\tau_1)h(\tau_2)R_{xx}(\tau_1-\tau_2-\tau)\; d\tau_1 d\tau_2$
	$R_{yy}(\tau) = R_{yy}(-\tau)$

Table 1 (Continued)

LINEAR SYSTEM RELATIONSHIPS (CONTINUED)

Spectral Density Relationships	$\Phi_{xy}(f) = H(-2\pi f)\Phi_{xx}(f)$
	$= H^*(2\pi f)\Phi_{xx}(f)$
	$\Phi_{yx}(f) = H(2\pi f)\Phi_{xx}(f)$
	$= H^*(-2\pi f)\Phi_{xx}(f)$
	$\Phi_{yy}(f) = \lvert H(2\pi f)\rvert^2\Phi_{xx}(f)$

*Denotes complex conjugate.

9 TRANSFER FUNCTION ESTIMATES AND COHERENCE

In the previous section the relationships between stochastic process properties of the input and output signals of a linear dynamic system are derived, given the properties of the linear system. Often in engineering practice we suspect that two observed stochastic signals are related by some linear system which we have not characterized. In this section we are concerned with the questions: Are two observed signals related by a linear dynamic system? And if so: What is the transfer function of that system?

Given any two time functions x(t) and y(t), we could compute their power and cross-spectral densities, and then by using equations of table 1 solve for an estimate $\hat{H}(w)$ of the unknown transfer function. This procedure gives

$$\hat{H}(w) = \Phi_{yx}(w/2\pi)/\Phi_{xx}(w/2\pi) \qquad w \geqq 0$$

and

$$\hat{H}(w) = \hat{H}^*(-w) \qquad w < 0 \tag{2-81}$$

If only the magnitude of the estimated transfer function were required, an alternative approach is

$$|\hat{H}(w)|^2 = \Phi_{yy}(w/2\pi)/\ \Phi_{xx}(w/2\pi) \tag{2-82}$$

or equivalently

$$\hat{H}(w) = \Phi_{yy}^{1/2}\ (w/2\pi)\ \Phi_{xx}^{-1/2}\ (w/2\pi) \tag{2-83}$$

As subsequently derived, equations (2-81) and (2-82) give the best possible estimates (in the mean-square-error sense) available from the data y(t) and x(t). However, additional calculations must be made to ascertain whether or not the signals x(t) and y(t) are indeed related by a linear system. These calculations are required since the quantities Φ_{yy} and Φ_{xx} are strictly positive and hence nonzero transfer functions estimates H(w) are obtained for any two signals x(t) and y(t) regardless of their connection to a single linear system. The signals need not even be measured simultaneously to give non-zero transfer function estimates. We shall define and compute a coherence function which plays a role analogous to that of the correlation coefficient introduced in section II.5 for random variables to characterize the extent of linear correlation between two time signals.

Recall from section II.5 that given data sets x_i and y_i an estimate $\hat{y}_i$ can be derived which is a linear function of x_i. The coefficients a and b of the functional relationship given in equation (2-84) are derived to minimize the mean-square-error.

$$\hat{y}_i = ax_i+b \tag{2-84}$$

As shown in section II.5 and Example 2-3, the coefficients a and b can be calculated from the covariance and the variances of the random variables x and y or they can be estimated from the data sets themselves. An analogous procedure is derived from estimating the linear dependence of two stochastic processes x(t) and y(t).

Given the stochastic signals x(t) and y(t), we assume that they are related by a linear system. Logically, an estimate $\hat{y}$ of y can be written as a convolution integral like that of equation (2-56). Explicitly we form the estimate

$$\hat{y}(t) = \int_{-\infty}^{\infty} h(\tau)\, x(t-\tau)\, d\tau \tag{2-85}$$

Since only the functions x(t) and y(t) are given, no prior knowledge of the system impulse response function $h(\tau)$ or its Fourier transform H(w) is available. These functions must be estimated from the given data. Some estimation error criterion must be formulated and the transfer functions chosen to minimize the error functional. The mean-square error functional leads most simply to calculable results. The estimation error is defined as

$$e(t) = y(t) - \hat{y}(t) \tag{2-86}$$

The autocovariance of the error is readily calculated as

$$\begin{aligned} R_{ee}(\tau) &= \mathcal{E}[e(t+\tau)\, e(t)] \\ &= \mathcal{E}\left\{\left[y(t+\tau) - \int_{-\infty}^{\infty} h(\tau_1)\, x(t+\tau-\tau_1)\, d\tau_1\right]\right. \\ &\quad \left.\cdot \left[y(\tau) - \int_{-\infty}^{\infty} h(\tau_2)\, x(t-\tau_2)\, d\tau_2\right]\right\} \\ &= R_{yy}(\tau) - \int_{-\infty}^{\infty} h(\tau_1)\left[R_{xy}(-\tau_1-\tau) + R_{xy}(\tau-\tau_1)\right] d\tau_1 \\ &\quad + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\tau_1)\, h(\tau_2)\, R_{xx}(\tau_1-\tau_2-\tau)\, d\tau_1\, d\tau_2 \end{aligned} \tag{2-87}$$

The mean-square error can be minimized by proper choice of the system impulse response h(t). We define F as those components of $R_{ee}(0)$ dependent upon h. We minimize $R_{ee}(0)$ by equivalently minimizing F.

$$F = -2\int_{-\infty}^{\infty} h(\tau_1)\, R_{xy}(-\tau_1)\, d\tau_1$$

$$+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\tau_1)\, h(\tau_2)\, R_{xx}(\tau_1-\tau_2)\, d\tau_1\, d\tau_2$$

$$=\int_{-\infty}^{\infty} h(\tau_1)\left[\int_{-\infty}^{\infty} h(\tau_2)R_{xx}(\tau_1-\tau_2)\, d\tau_2 - 2\, R_{xy}(-\tau_1)\right] d\tau_1 \qquad (2\text{-}88)$$

A variational method yields conditions which the minimizing function $h(\tau)$ must satisfy. Let $h^0(\tau)$ denote the minimizing solution and form

$$h(\tau) = h^0(\tau) + \varepsilon h^1(\tau)$$

where $h^1(\tau)$ is an arbitrary function and ε is a scalar parameter. Substituting for $h(\tau)$ into equation (2-88) and simplifying gives

$$F = F^0 + \varepsilon F^1 + \varepsilon^2 F^2 \qquad (2\text{-}89)$$

where

$$F^0 = \int_{-\infty}^{\infty} h^0(\tau_1)\left[\int_{-\infty}^{\infty} h^0(\tau_2)\, R_{xx}(\tau_1-\tau_2)\, d\tau_2 - 2\, R_{xy}(-\tau_1)\right] d\tau_1 \qquad (2\text{-}90)$$

$$F^1 = -2\int_{-\infty}^{\infty} h^1(\tau_1)\, R_{xy}(-\tau_1)\, d\tau_1$$

$$+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [h^1(\tau_1)\, h^0(\tau_2) + h^0(\tau_1)\, h^1(\tau_2)]\, R_{xx}(\tau_1 - \tau_2)\, d\tau_1\, d\tau_2$$

$$= 2\int_{-\infty}^{\infty} h^1(\tau_1)\left[\int_{-\infty}^{\infty} h^0(\tau_2)\, R_{xx}(\tau_1 - \tau_2)\, d\tau_2 - R_{xy}(-\tau_1)\right] d\tau_1 \qquad (2\text{-}91)$$

and

$$F^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h^1(\tau_1)\, h^1(\tau_2)\, R_{xx}(\tau_1 - \tau_2)\, d\tau_1\, d\tau_2 \qquad (2\text{-}92)$$

It can be shown that $F^2 \geqq 0$ for arbitrary $h^1(\tau)$.

The F^1 term dominates any variation of F with ε for ε sufficiently small. Since ε may assume both positive or negative values, a necessary condition that $h^0(\tau)$ be the minimizing solution is the requirement that

$$F^1 \equiv 0 \quad \text{for arbitrary } h^1(\tau) \qquad (2\text{-}93)$$

which in turn requires that

$$R_{xy}(-\tau) = \int_{-\infty}^{\infty} h^0(\tau_2)\, R_{xx}(\tau - \tau_2)\, d\tau_2 \qquad (2\text{-}94)$$

The integral equation (2-94) is one variant of the so-called Wiener-Hopf equation (reference 5, p. 305). Solution for a function $h^0(\tau_2)$ which satisfies equation (2-94) is most readily obtained by Fourier transforming that equation to obtain an equation for $H^0(w)$ which can be inverse transformed to yield a solution. Effecting the transform

$$F[R_{xy}(-\tau)] = \hat{H}^0(w)\, F[R_{xx}(\tau_1)] \tag{2-95}$$

denoting

$$G_{yx}(w) = F[R_{yx}(\tau)] = F[R_{xy}(-\tau)] \tag{2-96}$$

and

$$G_{xx}(w) = F[R_{xx}(\tau)] \tag{2-97}$$

gives the solution

$$\hat{H}^0(w) = G_{yx}(w)/G_{xx}(w) \tag{2-98}$$

and

$$h^0(\tau) = F^{-1}[\hat{H}^0(w)] \tag{2-99}$$

In general the impulse response function obtained by this technique is nonzero for $\tau \leq 0$. Thus, no physical system could be constructed having this impulse function since all realizable systems are non-anticipating (they cannot begin to respond to an input that has not yet been applied). A spectral factorization technique developed by Norbert Wiener (reference 7) does generate a solution of equation (2-94) which is implementable by casual systems. This approach is most readily used in the case that the functions $G_{yx}(w)$ and $G_{xx}(w)$ are expressed as rational polynominals in w. Since such polynominal expressions are not available from our analysis we discuss his technique no further.

Summarizing these results, if $G_{yx}(w) = F[R_{xy}(-\tau)]$ and $G_{xx}(w) = F[R_{xx}(\tau)]$, then the best estimate $\hat{y}(t)$ of y obtainable from the data x(t) is

$$y(t) = \int_{-\infty}^{\infty} h(\tau)\, x(t-\tau)\, d\tau \tag{2-100}$$

where

$$h(\tau) = F^{-1}[G_{yx}(w)/G_{xx}(w)]$$

The estimation error obtained in the minimization is $R_{yy}(o) + F^0$. We cally manipulate that expression by use of equations (2-90) and (2-94 a more useful result.

$$F^0 = -\int_{-\infty}^{\infty} h^0(\tau_1)\, R_{xy}(-\tau_1)\, d\tau_1 = -\frac{1}{2\pi}\int_{-\infty}^{\infty} H(w)G_{yx}(-w)dw$$

$$= -\frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{|G_{yx}(w)|^2}{G_{xx}(w)}\, dw$$

and consequently

$$E = R_{yy}(o) - \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{|G_{yx}(w)|^2}{G_{xx}(w)}\, dw$$

Recall from equation (2-94) that

$$G_{yy}(w) = F[R_{yy}(\tau)]$$

Inverse transforming and evaluating at $\tau = o$ yields

$$R_{yy}(o) = \frac{1}{2\pi}\int_{-\infty}^{\infty} G_{yy}(w)dw$$

and we can write

$$E = \frac{1}{2\pi}\int_{-\infty}^{\infty} \left[G_{yy}(w) - \frac{|G_{yx}(w)|^2}{G_{xx}(w)}\right] dw$$

or equivalently

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} G_{yy}(w) \left[1 - \frac{|G_{yx}(w)|^2}{G_{xx}(w)\ G_{yy}(w)}\right] dw \qquad (2\text{-}107)$$

This latter expression is particularly useful since it allows a definition of normalized error on a spectral basis. We define the coherence function of the estimate as

$$\gamma_{xy}^2(f) = \gamma_{xy}^2\left(\frac{w}{2\pi}\right) = \frac{|G_{yx}(w)|^2}{G_{xx}(w)\ G_{yy}(w)} = \frac{|\Phi_{yx}(f)|^2}{\Phi_{xx}(f)\ \Phi_{yy}(f)} \qquad (2\text{-}108)$$

Clearly

$$0 \leq \gamma_{xx}^2(f) \leq 1 \qquad (2\text{-}109)$$

The coherence function plays a role similar to the correlation coefficient. It characterizes the degree of the linear relationship between two signals such that as $\gamma_{xy}^2(f)$ approaches 1 the two functions are more highly correlated, while $\gamma_{xy}^2(f)$ equal to zero guarantees no linear dependence.

EXAMPLE 2-6

Consider a linear dynamic system with input x(t) and response y(t). Power spectral densities of the input and output signals have been measured as well as the input-output cross spectral density. An estimate of the system transfer function is to be obtained from these quantities. The coherence function is also computed and the results interpreted. The measured spectral densities are

$$G_{xx}(w) = 1$$

$$G_{yy}(w) = \frac{1}{1+(w/a)^2} + .01$$

and

$$G_{yx}(w) = \frac{1}{1 + j\ w/a}$$

From equation (2-98) the transfer function estimate is computed as

$$\hat{H}(w) = \frac{G_{yx}(w)}{G_{xx}(w)} = \frac{1}{1+jw/a}$$

The coherence function is computed by substitution for $G_{yx}(w)$, $G_{yy}(w)$, and $G_{xx}(w)$ in equation (2-108).

$$\gamma^2_{xy}(w/2\pi) = \frac{|1 + jw/a|^{-2}}{\frac{1}{1 + (w/a)^2} + .01} = \left\{1 + .01\,[1 + (w/a)^2]\right\}^{-1}$$

Observe that at low frequencies where $w < a$ we have $\gamma^2 \approx 1$. At these frequencies the estimate $\hat{H}(w)$ of the transfer function is very good. At frequencies $w > a$, γ^2 is less than one and it approaches zero as w becomes large. The estimate is not good at high frequencies. A block diagram description of a system having the properties given in this example is presented in figure 18.

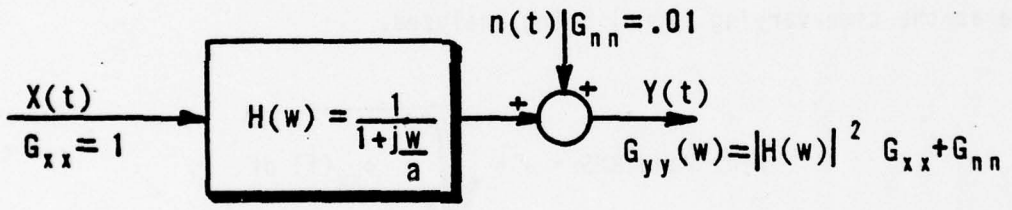


Figure 18. Linear System Transfer Function Identification for Example 2-6.

The estimated transfer function agrees identically with the actual system transfer function. The coherence function decreases from unity at higher frequencies, however, because the measurement noise n(t) becomes proportionately large with respect to the linear system output. At very high frequencies, the signal y(t) is dominated by n(t) and, consequently, y(t) is not accurately modeled by y(t) as obtained from equations (2-100) and (2-101). The coherence function deviation from unity indicates this fact.

Additional properties of the coherence function and its role in transfer function estimation are developed in section IV. We caution the reader that these techniques are applicable to stochastically excited systems. Transfer function estimates for linear systems with deterministic input signals are obtained by other techniques.

Other useful formats for the presentation of PSD information have been devised. These formats, all entailing integrals of PSD's over a given frequency band, are used most commonly for characterizing the performance of stochastically excited systems. Definitions for cumulative power (CUM PWR), wide band RMS, σ, inband RMS, and BACKWARD SUM are presented here for completeness. We comment briefly upon the interpretation of each term.

$$\text{CUM PWR } (f_H) = \int_0^{f_H} \Phi_{xx}(f)\, df \tag{2-110}$$

Cumulative power represents a signal's mean square value in frequency components from dc to a maximum frequency f_H. Wideband RMS represents the root-mean-square signal value contained in signal frequency components from dc to a maximum frequency. RMS is the level of a dc signal containing the same root-mean-square value as the time-varying signal being analyzed.

$$\text{Wide Band RMS} = \sigma = \sqrt{\int_0^{\infty} \Phi_{xx}(f)\, df} \tag{2-111}$$

Clearly,

$$\sigma = \sqrt{\text{COM PWR } (\infty)}$$

Inband RMS is simply a generalization of wideband RMS. The lower and upper frequency limit are specified. Inband RMS represents the signal power within the specified frequency band.

$$\text{In Band RMS } (f_L, f_H) = \sqrt{\int_{f_L}^{f_H} \Phi_{xx}(f)\, df} \tag{2-112}$$

Finally, BACKWARD SUM (f_L) represents the signal power attributable to frequency components above the specified frequency f_L.

$$\text{BACKWARD SUM } (f_L) = \int_{f_L}^{\infty} \Phi_{xx}(f)\, df$$

$$= \sigma^2 - \text{CUM PWR } (f_L) \qquad (2\text{-}113)$$

10 SUMMARY

The key concept developed in this chapter is that of a stochastic process representing signals and phenomena which cannot be accurately predicted and are of a random or statistical nature. These signals are characterized in terms of their mean, covariance and spectral density properties. Given these properties of a linear system input forcing function, and the system transfer function, one can calculate the corresponding properties of the system response. These calculations are important when calculating the response of an hypothesized system (no hardware exists) to a known stochastic environment and for the performance evaluation of such systems. The transfer function estimation techniques presented in section II.9 allow calculation of the linear dependence of two signals. These techniques can establish cause and effect relationship between seemingly unrelated signals.

SECTION III

COMPUTATIONAL ASPECTS OF ANALYSIS

1 INTRODUCTION

The primary emphasis of this section is the development of specific data analysis procedures based upon the theoretical results presented in the last section. In developing practical data analysis techniques, we shall prefer (or in some cases be forced) to make simplifying approximations. These approximations shall be examined in detail to determine their impact upon the interpretation of our results. The basic results of this section deal with development of algorithms suitable for use with finite duration, sample-data sequences. These numerical procedures are suitable for coding in a higher level language and execution in a minicomputer or a larger computer system. The Discrete Fourier Transform (DFT) is introduced as a sampled-data equivalent of the Fourier transform. The Fast Fourier Transform as a computationally efficient algorithm for calculating the DFT of a numerical sequence is explained. Equations for computing power and cross-power densities and the coherence function are presented at the conclusion of this section.

2 FINITE DURATION DATA INTERVALS

The power spectral density of a stochastic process $x(t)$ and the cross-spectral density of two processes $x(t)$ and $y(t)$ have been shown to be important characterizations of stochastic processes. These properties are sufficient for estimation of the linear dependence of two processes and for calculating the output power spectrum of a dynamic linear system excited by a stochastic process. In section II the power spectrum is shown to be the Fourier transform of the process autocorrelation function, equation (2-71). Thus the process PSD is readily calculated once the autocorrelation function has been determined. Theoretically, the autocorrelation function is determined by the second-order density function, equation (2-44), or, for ergodic processes, by the time-average correlation computed from an observed realization of the process, equation (2-51). Second-order density functions are generally not known a priori. Estimation of a process autocorrelation function is generally much simpler than estimation of the second-order density function and calculation of the autocorrelation function from it. In either approach, we are practically limited by a finite duration interval of data from which the estimates can be computed. We explicitly examine the effects of the finite duration data interval in the

calculation of a power spectral density as the Fourier transform of a time average autocorrelation function estimate.

The finite duration interval has two immediate consequences in the interpretation of our results. Limiting the data interval to a finite duration makes the autocorrelation function a random variable with a nonzero covariance. The finite interval also distorts the estimated spectrum of $R_{xx}(\tau)$.

The randomness introduced by using limited data is best described by the following observations. Define

$$R_{x_T x_T}(\tau) \triangleq \frac{1}{T}\int_{-\infty}^{\infty} x_T(t)\, x_T(t+\tau)\, dt \tag{3-1}$$

where

$$x_T(t) = \begin{cases} x(t), & 0 \leq t \leq \tau \\ 0, \text{ otherwise} \end{cases} \tag{3-2}$$

Note that $R_{xTxT}(\tau)$ is certainly a random variable since it is a function of the stochastic processes $x(t)$ and $x(t+\tau)$. Under a general set of assumptions (reference 4, section 9) the covariance of $R_{xTxT}(\tau)$ approaches zero as T approaches infinity. For engineering purposes we can treat the estimated correlation function just like any other deterministic function of τ. The finite duration interval, however restricts T to finite values. Hence, the covariance of the random variable $R_{xTxT}(\tau)$ remains non-zero. The finite covariance of the autocorrelation function estimate has the following practical interpretation. If we were to compute $R_{xTxT}(\tau)$ from different data intervals (or from different members of the time ensemble) of the same ergodic stochastic process, we would calculate different functions $R_{xTxT}(\tau)$ for each. The mean or average value of the different calculations is $R_{xx}(\tau)$; the various values are dispersed about the mean with a covariance which (in principle, at least) is calculable from the density functions of the underlying probability space. A PSD calculated by Fourier transforming $R_{xTxT}(\tau)$ necessarily has a corresponding covariance or uncertainty. Bounds on the PSD uncertainty are derived in section III.4.

Calculation of a PSD estimate from a finite duration data interval is also claimed to have spectral distortion. By direct calculation

$$G_{X_T X_T}(w) = F\left[R_{X_T X_T}(\tau)\right] = \frac{1}{T}\,|X_T(w)|^2 \tag{3-3}$$

where

$$X_T(w) = F[x_T(t)] \tag{3-4}$$

But clearly we can write $x_T(t)$ as the product of the untruncated stochastic process $x(t)$ and a window function $g(t)$ which is unity for t within the available data time frame and zero elsewhere. Observe that explicit calculation of $X_T(w)$ using this expression for $x_T(t)$ gives:

$$X\ (w) = \int_{-\infty}^{\infty} x(t)\ g(t)\ \exp\ (-jwt)\ dt \tag{3-5}$$

but

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(w_1)\ \exp\ (jw_1 t)dw_1 \tag{3-6}$$

so by substitution for x(t) and an interchange in the order of integration

$$X_T(w) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} X(w_1)\ \exp\ (jw_1 t)\ g(t)\ \exp(-jwt)\ dw_1 dt$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} X(w_1)\ \left[g(t)\ \exp\ [-j(w_1-w)\ t]\ dt\right] dw_1$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} X(w_1)\ G\ (w-w_1)\ dw_1 \tag{3-7}$$

where

$$G(w) = \int_{-\infty}^{\infty} g(t) \exp(-jwt)\, dt \qquad (3\text{-}8)$$

Equation (3-7) is interpreted as a complex convolution of the functions X(w) and G(w) (compare with the convolution of time functions given in equation (2-56)). We see that $X_T(w)$ represents a complex weighted average of the values $X(w_1)$ in a neighborhood of frequencies about w. The weighting function G(w) is simply the Fourier transform of the data window function. Generally |G(w)| has largest magnitude for $w \approx o$ and its magnitude approaches zero as w becomes large. Typically a tradeoff exists between the main lobe width and the rate at which |G(w)| rolls-off with increasing frequency. A sharp main lobe is desired so that $X_T(w)$ closely approximates X(w). A fast rolloff is required to prevent components $X(w_1)$ from significantly contributing to $X_T(w)$ whenever $w_1 \neq w$. Examples of various window functions and their corresponding transform magnitudes are plotted in figure 19.

The spectrum of $X_T(w)$ is a distorted representation of the original signal spectrum. This distortion is most readily seen for the case wherein X(w) has discrete frequency singularities, as for example, does $x(t) = \cos w_o t$. The actual spectrum |X(w)| and the truncated signal spectrum $|X_T(w)|$ are plotted in figure 20. Notice that $|X_T(w)|$ has non-zero values over frequency bands where |X(w)| is zero. This property of $|X_T(w)|$ is due to a <u>leakage phenomenon</u> associated with all finite duration window functions. The primary limitation of data windows is that they smooth the measured spectra, introduce leakage components, and limit the resolution of the final spectral estimate. These effects are minimized by a choice of T sufficiently large that these attendant degradations are inconsequential. Generally this requirement means that $T > 2\pi/\Delta f$ where Δf is any frequency difference we wish to resolve. The confidence of our estimate will also require larger T for better estimates. These relationships are presented in Section III.4.

The cosine-taper data window shown in figure 19d is employed in the AFWL SIGNAL data analysis program. This window function provides good sidelobe rejection while introducing minimum distortion to the original data signal. A dominant effect of the distortion is to reduce the variance of the windowed data with respect to the original unwindowed signal (ref. 2, p. 323). The windowed data

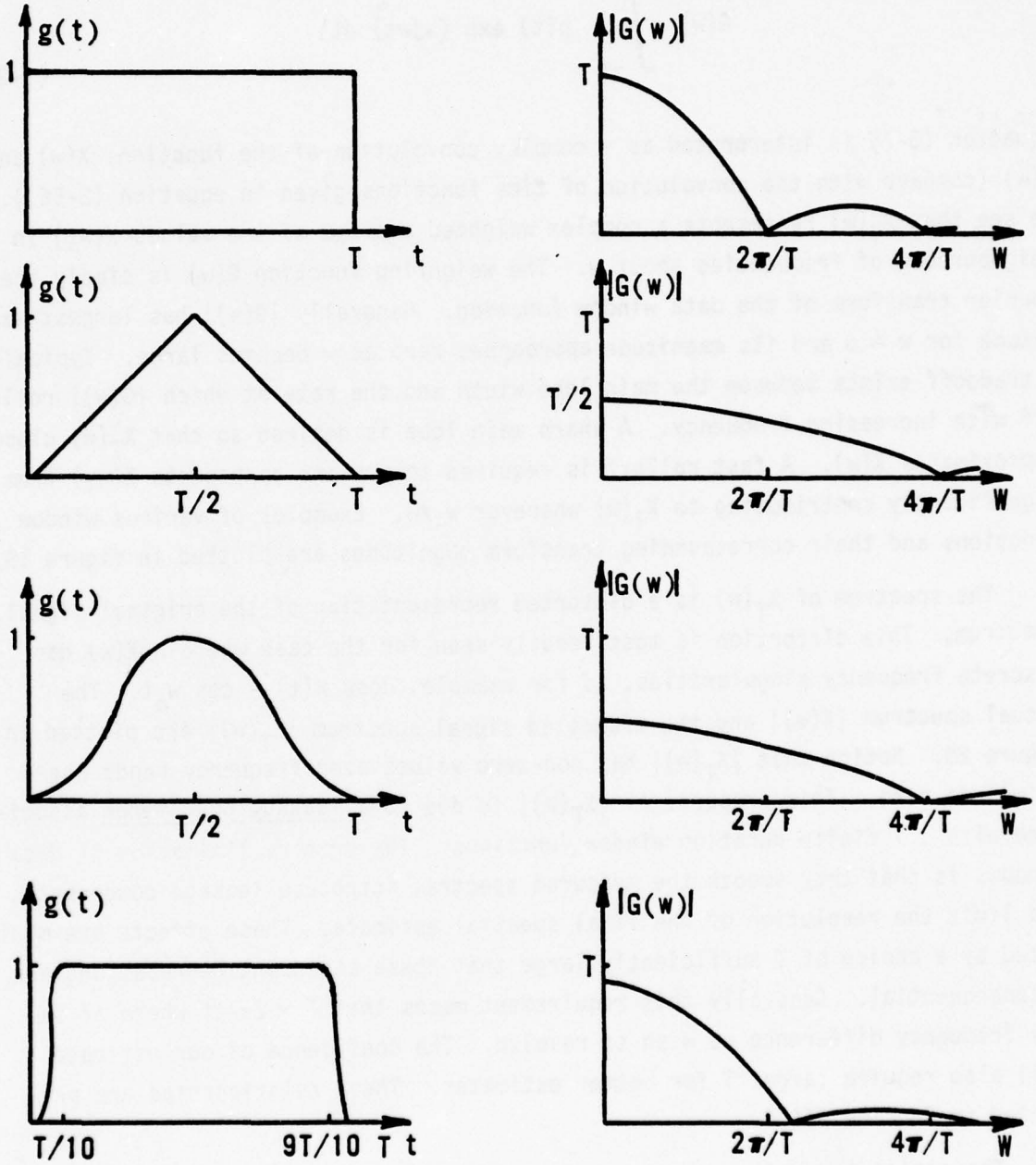


Figure 1. Data Windows and Their Transform Magnitudes (a) Boxcar, (b) Triangular, (c) Hanning, (d) Cosine Taper.

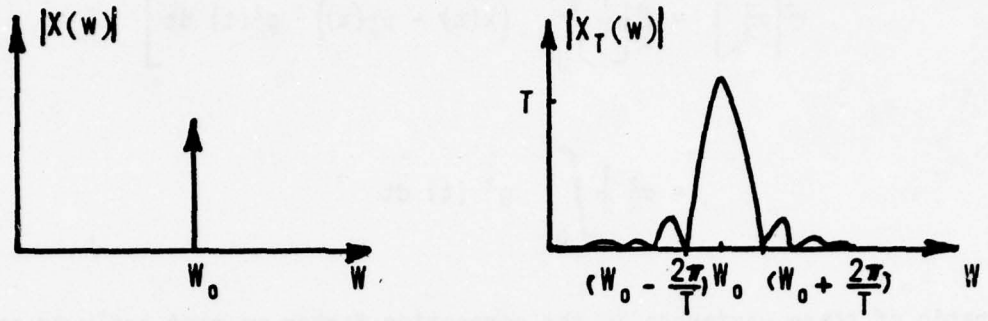


Figure 20. Spectral Leakage of Data Windows.

must be multiplied by a "correction factor" which increases the output data variation to its original, unwindowed value. The approximate correction factor is given by the ratio of the window function 'powers.' A window power is simply the area under the curve of the window function squared.

$$CF = \frac{\int_0^T g^2_{Boxcar}(t)\ dt}{\int_0^T g^2_{cosine}(t)\ dt} = \frac{1}{0.875} \tag{3-9}$$

This correction factor is precisely correct for random data as can be seen from the following calculations:

$$\mathcal{E}\left[\sigma^2_{X_T}\right] = \mathcal{E}\left[\frac{1}{T}\int_0^T \left(x(t) - \mu_x(t)\right)^2 dt\right]$$

$$= \sigma_x^2 \tag{3-10}$$

while for windowed data

$$\mathcal{E}\left[\sigma^2_{X_w}\right] = \mathcal{E}\left[\frac{1}{T}\int_0^T \left(x(t) - \mu_x(t)\right)^2 g^2(t)\, dt\right]$$

$$= \sigma^2_X \frac{1}{T}\int_0^T g^2(t)\, dt \qquad (3\text{-}11)$$

The ratio of these variances is the correction factor we must apply to approximately boost the variance of the windowed data to match the original signal variance.

$$CF = \frac{T}{\int_0^T g^2(t)\, dt} \qquad (3\text{-}12)$$

An alternant method that can be used to calculate the approximate correction factor is to actually calculate the sample variances of data both before and after multiplication by the window function. The approximate correction factor is then simply the ratio of these results. Koenigsberg (reference 8, p. 80) gives an excellent discussion of this procedure. His empirically founded observations provide useful interpretations regarding the systematicness of the data when the correction factors as calculated above differ greatly from that derived from equation (3-12).

The correction factor is a power boost multiplier for signal variances; hence, it is also the appropriate correction factor for PSD calculations. The complete PSD formulas are presented in section III.3.

3 SAMPLED DATA

The second most important aspect of practical stochastic data analysis procedures deals with the sampled-data nature of signals processed by numerical algorithms on a digital computer. The equations presented thus far have assumed that values of the time function x(t) are known for any t within the analysis interval [0,T]. Practically, this is the case only when the signals are specified analytically in a theoretical deviation or when the data analysis equations are implemented with analog, rather than digital, circuits. Prior to the widespread use of digital computers, analog bandpass filter techniques were employed

to estimate spectral densities. (Recall the PSD discussion in section II.8.) Digital numerical procedures make feasible and practical many new analysis techniques. Consequently, digitization of analog data signals and the numerical evaluation of our analysis equations requires careful examination as to the impact of these operations on our overall results. We shall find that data digitization introduces distortions of the estimated PSD just as did the finite duration data window.

Data digitization or sampling typically is a process whereby a continuous signal x(t) is converted to a sequence of digital words represented in a binary number system common to modern digital computers. Signal values corresponding to the digital sequence are obtained from the original signal at fixed time intervals ΔT. The sampling frequency or sampling rate is $f_s = \Delta T^{-1}$; ΔT is the sampling interval or period. A pictorial representation of the sampling process is presented in figure 21.

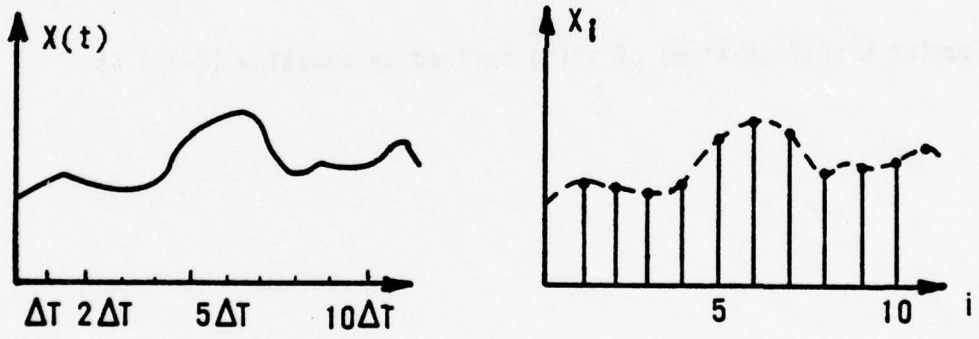


Figure 21. The Sampling Process for Signal Digitization.

Assume that the sample points $x_i = x\ (i\ \Delta T)$ for integer i are known rather than the values of the function x(t) for an arbitrary t. We wish to use these sample points to determine Fourier transform spectral properties of the unsampled time signal x(t). We observe that one possible convention for reducing the original function x(t) to a time function completely characterized by the values

x_i is to define the sampled-data time function $x^*(t)$ as the multiple of $x(t)$ with a time sequence of impulse functions spaced ΔT seconds apart.* The sequence of impulse functions is called an impulse train and is denoted $\delta_T(t)$.

$$\delta_T(t) = \sum_{i=-\infty}^{\infty} \delta(t-i\Delta T) \tag{3-13}$$

Performing explicitly the multiplication of the signal by the impulse train gives

$$x^*(t) = x(t)\ \delta_T(t)$$

$$= \sum_{i=-\infty}^{\infty} x(i\Delta T)\ \delta(t - i\Delta T) \tag{3-14}$$

The Fourier transform $X^*(w)$ of $x^*(t)$ defined by equation (2-57) is

*Recall that the impulse function (Dirac delta function) is defined by the following properties:

a. $\int_{-\varepsilon}^{\varepsilon} \delta(t)\, dt = 1$ for arbitrarily small $\varepsilon>0$

b. $\int_{\tau+\varepsilon}^{\tau+\varepsilon} f(t)\ \ \delta(t-\tau)\, dt = f(\tau)$ if f is continuous at τ

c. $\delta(t) = 0$ $(t>0$ or $t<0)$ (i.e., $t \neq o)$

$$X^*(w) = \int_{-\infty}^{\infty} x^*(t) \exp(-jwt)\, dt$$

$$X^*(w) = \sum_{i=-\infty}^{\infty} x_i \exp(-jw\, \Delta T i) \tag{3-15}$$

Certainly we could omit the intervening steps and simply define a <u>sample-data Fourier transform</u> as

$$X^*(w) = \sum_{i=-\infty}^{\infty} x(i\Delta T) \exp(-jw\, i\Delta T) \tag{3-16}$$

The time domain multiplication of the original signal $x(t)$ by the impulse train requires that the Fourier transform of the resulting sampled-data sequence be the convolution of the original data transform with the transform of the impulse train. The Fourier transform is readily seen to be a sequence of singularities at frequencies $w = n\frac{2\pi}{\Delta T}$, $i = 0, \pm 1, \pm 2, \ldots$. This property readily follows from the fact that $\delta_T(t)$ is periodic with periodicity ΔT. Also

$$F[\delta_T(t)] = \sum_{i=-\infty}^{\infty} \exp(-jwi\Delta T)$$

$$= \infty \text{ for } w = \frac{k2\pi}{\Delta T} \tag{3-17}$$

We shall show that the sampled-data Fourier transform is related to the original signal transform by the equation

$$X^*(w) = \frac{1}{\Delta T} \sum_{n=-\infty}^{\infty} X\left(w + \frac{2\pi}{\Delta T} n\right) \tag{3-18}$$

where

$$X(w) = F[x(t)]$$

Typical plots of $|X^*(w)|$ given $|X(w)|$ are plotted in figure 22. The extremely

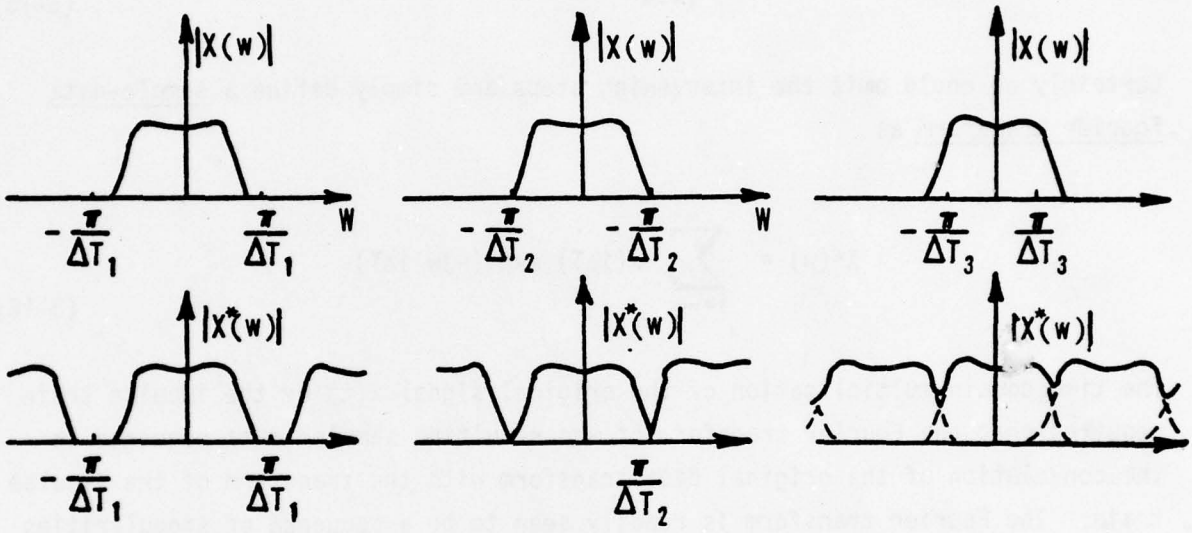


Figure 22. Spectral Folding of Sampled Signals.

important result is that frequency components in X(w) above the Nyquist frequency $w = \pi/\Delta T$ are folded-down to frequencies within the band $|w| \leq \pi/\Delta T$. Thus these higher frequency components, mimic, or alias spectral components in the lower frequency band. A time domain interpretation of the aliasing phenomena is shown in figure 23. Observe the spectral distortion which occurs in the case that the sample rate is too low. Distortion is minimized with sample frequencies at least twice the signal bandwidth. Signals which are bandlimited to the Nyquist frequency can be uniquely represented by their sample sequence x_i. Unfortunately, perfectly bandlimited signals are seldom encountered in practice. Many practical signals are approximately bandlimited, however. Those signals not sufficiently bandlimited for the desired sample rate can be appropriately filtered by anti-aliasing filters to minimize their spectral content beyond the Nyquist frequency thereby minimizing aliasing distortion. Anti-aliasing filtering must be accomplished prior to digitization.

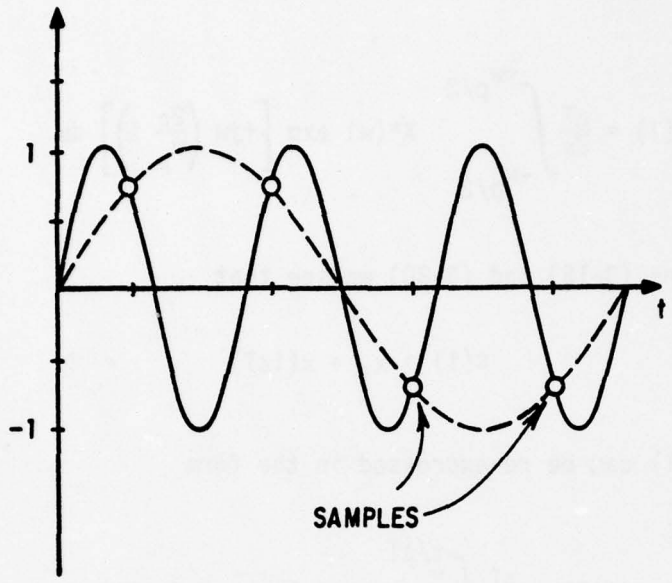


Figure 23. A Time Domain Example of Frequency Aliasing.

The important relationship of equation (3-18) is proven in the following manner. First note that X*(w) as expressed in equation (3-15) is periodic in w with periodicity $w_p = 2\pi/\Delta T$. This periodicity is a consequence of the periodicity of the complex exponential function.

$$\exp\left[-j\left(\frac{2\pi}{\Delta T}n+w\right)\Delta Ti\right] = \exp(-j2\pi ni)\exp(-j w\Delta Ti)$$

$$= \exp(-jw\Delta Ti) \qquad (3\text{-}19)$$

since $\exp(-j2\pi ni) = 1$ for all integers n and i. The periodic nature of X*(w) permits it to be represented by a sum of sinusoids (Fourier series) of the form

$$X^*(w) = \sum_{i=-\infty}^{\infty} c(i)\exp\left[-j\, w\left(\frac{2\pi}{w_p}\right)i\right]$$

$$= \sum_{i=-\infty}^{\infty} c(i)\exp(-jw\Delta Ti) \qquad (3\text{-}20)$$

where the Fourier coefficients c(i) are given by

$$c(i) = \frac{\Delta T}{2\pi}\int_{-w_{p/2}}^{+w_{p/2}} X^*(w)\ \exp\left[+jw\left(\frac{2\pi}{w_p}\,i\right)\right] dw \tag{3-21}$$

Comparing equations (3-15) and (3-20) we see that

$$c(i) = x_i = x(i\Delta T) \tag{3-22}$$

and equation (3-21) can be re-expressed in the form

$$x(i\Delta T) = \frac{\Delta T}{2\pi}\int_{-\pi/\Delta T}^{\pi/\Delta T} X^*(w)\ \exp\ (+jw\Delta T i)\ dw \tag{3-23}$$

Standard Fourier analysis of the original signal expressed as in inverse transform is

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(w)\ \exp\ (jwt)\ dw$$

$$= \frac{1}{2\pi}\sum_{m=-\infty}^{\infty}\int_{\frac{(2m-1)\pi}{\Delta T}}^{\frac{(2m+1)\pi}{\Delta T}} X(w)\ \exp\ (+jwt)dw$$

$$= \frac{1}{2\pi}\int_{-\frac{\pi}{\Delta T}}^{\frac{\pi}{\Delta T}}\sum_{m=-\infty}^{\infty} X\left(w + \frac{2\pi}{\Delta T}\,m\right)\exp\left(jwt+jm\frac{2\pi}{\Delta T}\,t\right) dw \tag{3-24}$$

Restricting attention to values of $t = i\Delta T$ in the above equation and making the simplification in the exponential that this allows gives

$$x(i\Delta T) = \frac{\Delta T}{2\pi} \int_{-\frac{\pi}{\Delta T}}^{+\frac{\pi}{\Delta T}} \left[\frac{1}{\Delta T} \sum_{m=-\infty}^{\infty} X\left(w + m\frac{2\pi}{\Delta T}\right) \right] \exp\,(jw\Delta Ti)\, dw \tag{3-25}$$

Comparing equations (3-23) and (3-25) and the uniqueness of the Fourier transform gives the identity

$$X^*(w) = \frac{1}{\Delta T} \sum_{m=-\infty}^{\infty} X\left(w + \frac{2\pi}{\Delta T}\, m\right) \tag{3-26}$$

as was to be proven.

Up until this point we have assumed that an infinite duration of data is available. The practical considerations of a finite duration data interval discussed in section III.2 also hold for sampled-data. This finite data restriction reduces equation (3-15) to the form

$$X^*_T(w) = \sum_{i=0}^{N-1} x_i \exp(-jw\Delta Ti) \tag{3-27}$$

The above representation of the Fourier transform of the signal $x(t)$ is simply computed for a given w as the weighted sum of exponentials. We shall show that $X^*_T(w)$ need be evaluated for only N equally spaced values of w in order to uniquely specify the sampled sequence x_i. Then we shall show how these N values may be efficiently calculated.

The discrete Fourier transform (DFT) is defined simply as the Fourier transform of a finite length sample sequence evaluated at the discrete frequencies $w = n\frac{2\pi}{\Delta T}$, $n = o, 1, 2 \ldots N - 1$. We now show the correspondence between the DFT and the transform $X(w)$ of the original unsampled signal.

Consider the windowed time function x(t) defined on the interval [0,T) and elsewhere zero. Let x_i, i = 0, 1, . . . ,N-1 be the corresponding sample sequence taken at times t = iΔT, where $\Delta T = T/N$. Since x(t) is zero outside the interval [0,T), we may assign values to a function $\hat{x}(t)$ such that $\hat{x}(t) = x(t)$ for t in [0,T) and in such a manner that allows us to most easily reconstruct x(t) from $\hat{x}(t)$ and its transform. One potentially useful assignment is to define $\hat{x}(t)$ as the periodic repetition of x(t) as shown in figure 24.

$$\hat{x}(t) = \sum_{n=-\infty}^{\infty} x(t+nT) \tag{3-28}$$

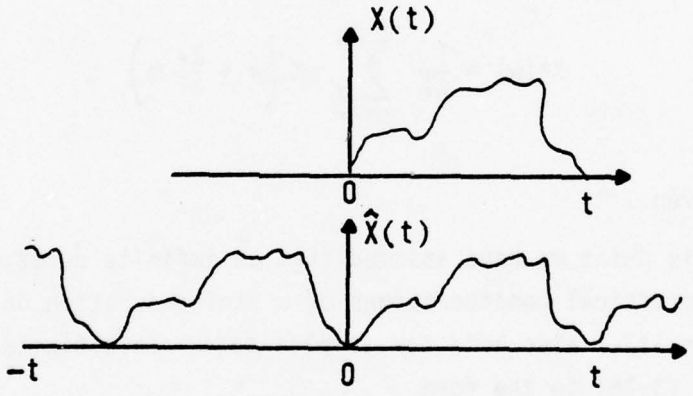


Figure 24. Periodic Continuation of a Finite Duration Signal.

The advantage of this choice is that because of its periodicity, it can be represented by a Fourier series at frequencies $w_p = \frac{2\pi}{T}$ and its harmonics. Further, if x(t) is truly bandlimited at $w_L = \frac{\pi}{\Delta T}$ then only $W_L/w_p = N/2$ components need be calculated to uniquely characterize x(t). We shall show that knowledge of N DFT points is sufficient to characterize x_i. Knowledge of transform values at other frequencies is not required and in fact these values may be deduced from the discrete frequency transform values that are actually evaluated.

In terms of a complex Fourier series, x(t) can be written

$$\hat{x}(t) = \sum_{k=-\infty}^{\infty} X_p^1(k) \exp\left(+j\frac{2\pi}{T} kt\right) \tag{3-29}$$

or when t is restricted to values $i\Delta T$

$$\hat{x}(i\Delta t) = \sum_{k=-\infty}^{\infty} X_p^1(k) \exp\left(j\frac{2\pi}{T} k\Delta T i\right)$$

$$= \sum_{k=-\infty}^{\infty} X_p^1(k) \exp\left(j\frac{2\pi}{N} ki\right) \tag{3-30}$$

since $T/\Delta T = N$. The exponential term is periodic in k with periodicity N so that equation (3-30) may be written as a finite sum

$$\hat{x}(i\Delta T) = \frac{1}{N} \sum_{k=0}^{N-1} X_p(k) \exp\left(j\frac{2\pi}{N} ki\right) \tag{3-31}$$

The transform $X_p(k)$ is seen to be

$$X_p(k) = N \sum_{\ell=-\infty}^{\infty} X_p^1(k+\ell n) \tag{3-32}$$

The transform $X_p(k)$ is most readily evaluated from equation (3-31) rather than by use of equation (3-32). Multiplying both sides of equation (3-31) by $\exp\left(-j\frac{2\pi}{N} i\ell\right)$ and summing over the index i gives

$$\sum_{i=0}^{N-1} x(i\Delta T)\ \exp\ \left(-j\frac{2\pi}{N}\ i\ell\right) = \sum_{i=0}^{N-1} \frac{1}{N} \sum_{k=0}^{N-1} X_p(k)\ \exp\ \left[j\frac{2\pi}{N}(k-\ell)\right]$$

$$= \sum_{k=0}^{N-1} X_p(k) \left[\frac{1}{N} \sum_{i=0}^{N-1} \exp\ j\frac{2\pi}{N}(k-\ell)i\right] \qquad (3\text{-}33)$$

The term within the brackets is recognized as a geometric series which is summable in closed form. It equals $\frac{1}{}$ whenever $(k - \ell)$ is an integer multiple of N and is zero otherwise. Hence we have shown the transform pair

$$x(i\Delta T) \mathrel{<=>} X_p(k) \qquad (3\text{-}34)$$

The transform relationships are summarized by the equations

$$X_p(k) = \sum_{i=0}^{N-1} x(i\Delta T)\ \exp\ \left(-j\ \frac{2\pi}{N}\ ik\right) \qquad (3\text{-}35)$$

and

$$x(i\Delta T) = \frac{1}{N} \sum_{k=0}^{N-1} X_p(k)\ \exp\ \left(j\ \frac{2\pi}{N}\ ki\right) \qquad (3\text{-}36)$$

Equations (3-35) and (3-36) are the DFT and inverse DFT, respectively, of the sampled-data sequence $x_i = x(i\Delta T)$. By reconstructing the steps leading to equation (3-35), the relationship between $X_p(k)$ and $X(w)$, the Fourier transform of the original data signal, is established. These steps are summarized in table 2. Observe that the order in which Steps 1 and 2 are executed may be reversed without changing the result. The important conclusion to be drawn from table 2, given the assumptions that

Table 2

THE RELATIONSHIP OF DISCRETE FOURIER TRANSFORMS TO THE ORIGINAL SIGNAL TRANSFORM

Step	Process	Input	Output	Transform Relationship
1	Sampling	$x(t)$	$x^*(t) = x(t)\cdot\delta_T(t)$	$X^*(w) = \frac{1}{\Delta T}\sum_{n=-\infty}^{\infty} X\left(w+\frac{2\pi}{\Delta T}n\right)$
2	Windowing	$x^*(t)$	$x_T^*(t) = g(t)\cdot x^*(t)$	$X_T^* = X^*(w)* G(w)$
				$= \frac{1}{\Delta T}\sum_{n=-\infty}^{\infty}\left[X^*\left(w+\frac{2\pi}{N}n\right)\ G(w)\right]$
3	Discrete Fourier Transform	x_i	$x_i = x_T^*(i\Delta T)$	$X_p(k) = X^*(w)\ \|_{w=k\frac{2\pi}{\Delta T}}$

(1) the sampling rate is sufficiently fast that

$$X^*(w) \approx \frac{1}{\Delta T} X(w)\ ,|w| < \frac{\pi}{\Delta T}$$

and

(2) the data analysis interval is sufficiently long that the frequency distortion induced by windowing is negligible (even though a correction factor may be required).

is that

$$X_p(k) \approx \frac{X(w)}{\Delta T\sqrt{CF}}\ \Big|_{w\,=\,k\,\frac{2\pi}{T}} \tag{3-37}$$

or

$$X(w)\ \Big|_{w=k\,\frac{2\pi}{T}} \approx \Delta T\ X_p(k)\ \sqrt{CF} \tag{3-38}$$

where $X_p(k)$ is the DFT of data sequence. It is defined by equation (3-35). This method for computing X(w) and the definitions of power and cross spectral densities, transfer function estimates and the coherence function establish the approximations

PSD
$$\Phi_{xx}\left(\frac{k}{T}\right) = \frac{2}{T}\,|X_T(w)|^2\,\Big|_{w=\frac{2\pi}{T}k} \approx \frac{2\,CF}{N^2}\,|X_p(k)|^2\,T \tag{3-39}$$

CSD
$$\Phi_{xy}\left(\frac{K}{T}\right) = 2\,\frac{X_T(w)\,Y_T^*(w)}{T}\Big|_{w=k\,\frac{2\pi}{T}} = 2\frac{CF}{N^2}\,X_p(k)Y_p(-k)T \tag{3-40}$$

TRANSFER FUNCTION ESTIMATES

(Method 1)
$$\hat{H}\left(k\,\frac{2\pi}{T}\right) = \Phi_{yx}\left(\frac{k}{T}\right)/\,\Phi_{xx}\left(\frac{k}{T}\right) = Y_p(k)/\,X_p(k) \tag{3-41}$$

(Method 2)
$$\left|\hat{H}\left(k\,\frac{2\pi}{T}\right)\right| = \Phi_{yy}^{1/2}\left(\frac{k}{T}\right)\Phi_{xx}^{-1/2}\left(\frac{k}{T}\right) = \left|Y_p(k)/\,X_p(k)\right| \tag{3-42}$$

Coherence Function

$$\gamma_{xy}^2\left(\frac{k}{T}\right) = \frac{\left|\Phi_{xy}\left(\frac{k}{T}\right)\right|^2}{\Phi_{xx}\left(\frac{k}{T}\right)\Phi_{yy}\left(\frac{k}{T}\right)} \equiv 1 \quad \text{(see note)} \tag{3-43}$$

Note: The coherence function as evaluated here is a highly biased estimate of the actual true value of the coherence. This difficulty is removed by application of frequency averaging techniques developed in section III.4.

Recall that the above equations represent estimated quantities and the above expressions are "raw" estimates. Averaging techniques for reducing the variance of the estimates are developed in section III.4.

The use of equations (3-39) through (3-43) became particularly attractive for digital computer evaluation of signal properties with the introduction of the Fast Fourier Transform, FFT, by Cooley and Tukey (reference 9) in 1965. The FFT is a computational efficient algorithm for evaluation of the DFT, equation (3-35).

The FFT algorithm depends upon two key aspects. One aspect is that the computational time required to evaluate $X_p(k)$, $k = 0, 1, \ldots N-1$ is dominated by the time required to compute a complex multiplication. An N point sequence requires N^2 multiplications, a computational burden which is prohibitive for large N. The FFT 'trades' complex multiplications for an increased number of computationally less demanding complex additions to achieve a computational complexity comparable to $N \log_2 N$ complex multiplications. The computational savings is the ratio

$$\frac{N^2}{N \log_2 N} = \frac{N}{\log_2 N} \qquad (3\text{-}44)$$

The computational savings ratio is tabulated in table 3 for typically used values of N.

Table 3

THE COMPUTATIONAL SAVINGS RATIO OF THE FAST FOURIER TRANSFORM

N	$N/\log_2 N$
512	57
1024	102
4096	341
16384	1170

For example, the numerical evaluation of a 512 point DFT by straight forward evaluation of equation (3-35) would require 57 times as much computer time as would the FFT to evaluate the same 512 point sequence. A second aspect of the DFT formula which makes the FFT possible is the periodicity of the complex exponential.

$$\exp\left[j \frac{2\pi}{N} (i+\ell n)\right] = \exp\left(j\frac{2\pi}{N} i\right) \tag{3-45}$$

These aspects of computing DFT's lead to computational advantage as follows: First we observe that the mathematical expression E = ab + ac + ad can be evaluated explicitly as written to require three multiplications and two additions. Alternatively, E = a(b+c+d) evaluated explicitly requires two additions but now requires only a single multiplication to obtain an equivalent result. This concept is similarly applied toward evaluation of the DFT.

We separate a N point sequence, assuming N is an even number, into two N/2 point sequence x_{2i} and $x_{2\ell+1}$. One sequence has been formed from the points having even indices and the other from the points having odd indices. Applying equation (3-35)

$$\text{DFT}\,(x_i) = X_p(k) = \sum_{i=0}^{N-1} x_i \exp\left(-j \frac{2\pi}{N} ik\right)$$

$$= \sum_{i=0}^{N/2-1} \left[x_{2i} \exp\left(-j \frac{2\pi}{N} 2ik\right) + x_{2i+1} \exp\left(-j \frac{2\pi}{N} \overline{2i+1}k\right)\right]$$

$$= \sum_{i=0}^{N/2-1} \left[x_{2i}+x_{2i+1} \exp\left(-j \frac{2\pi k}{N}\right)\right] \exp\left(-j \frac{2\pi}{N/2} ik\right)$$

$$= \text{DFT}\left(x_{2i}\right) + \exp\left(-j \frac{2\pi}{N} k\right) \text{DFT}_i\left(x_{2i+1}\right) \tag{3-46}$$

Thus the original N point transform has been changed to an equivalent problem of evaluating two N/2 point sequences and linearily combining the results. The computational saving accrues because the computation burden is proportional to the sequence length squared. The N^2 multiplications required for direct evaluation of the DFT is reduced to $2(N/2)^2 + N \approx N^2/2$ multiplications.

A secondary advantage of the FFT technique is that one evaluation of the complex exponential can be used for the evaluation of both the two smaller sequences. Thus, fewer evaluations of the complex exponential are required.

The process of dividing evaluation of an DFT into the evaluation of two DFT's having half as many points, as done in equation (3-46), can be repeated as long as the number of points in the DFT actually evaluated is even. Cooley and Tukey (reference 9) developed the algorithm for special case that $N = 2^m$, m an integer. In this case, N multiplications are performed for each of the m steps in which DFT's are expressed as the linear combinations of two shorter DFT's. The total number of multiplications required by this method is

$$Nm = n\log_2 N \tag{3-47}$$

compared with N^2 multiplications required for direct evaluation of the DFT given by equation (3-35).

Other methods applicable to cases in which N is not a power of 2 have also been developed (references 10, 11, and 12). These methods involve the decomposition of N into its component prime factors. Each of these methods is equally applicable to evaluation of the inverse DFT, equation (3-36).

3 CONFIDENCE INTERVALS

In section II.5 we showed that practical considerations of random data analysis prevented us from precisely calculating the statistics of an observed process. Rather, sample statistics can be calculated which approximate the desired process statistics. In this section we shall derive a measure of the error or quality of our estimates.

Generally, the measurement error by which a sample statistic differs from the true quantity is a random variable since it is the function of random variables. Consequently, mean values, moments and probability density functions characterize the error random variable. This aspect of sample statistics is illustrated with an example.

Given independent random variables x_i with means μ and variances σ^2, we make N observations. From these observations we wish to make an estimate μ of the process mean. A logical choice is to simply average the observations to form the estimate.

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{3-48}$$

Suppose now that we were to repeat our experiment. We would expect to make different observations x_i. Consequently, the estimate $\hat{\mu}$ we would compute with this second set of observations would differ from our first estimate. In fact, if we were to repeat our experiment infinitely often and compute the mean estimate for each case, we would find that the random variable $\hat{\mu}$ has a mean μ and variance σ^2/N. These quantities are readily calculated. Taking expectations of both sides of equation (3-48) establishes the mean value of $\hat{\mu}$.

$$\hat{\mu} = \mathcal{E}\left\{\frac{1}{N}\sum_{i=1}^{N} x_i\right\} = \frac{1}{N}\sum_{i=1}^{N}\mathcal{E}[x_i] = \mu \tag{3-49}$$

Similarly

$$\sigma^2_{\hat{\mu}} = \mathcal{E}[(\hat{\mu}-\mu)^2] = \mathcal{E}\left\{\left[\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)\right]^2\right\}$$

$$= \frac{\sigma^2}{N} \tag{3-50}$$

The uncertainty of our error depends upon the probability of the underlying process. Observe that for a very large N, the mean estimate becomes a very good one since the error variance decreases as N^{-1}.

The estimate error is a random variable. The question of exactly how much in error our particular estimate is cannot be answered. In many cases, however, we can compute the probability that the error exceeds a specified bound. Both the probability and bound must be specified to characterize the 'goodness' of our estimate.

These concepts are applied to the mean estimation problem developed in equations (3-48) through (3-50). We assume that the random variables x_i have normal distributions. Then the estimated mean is also normally distributed and has a PDF completely characterized by its mean and covariance given in equations (3-49) and (3-50).

$$p(\hat{\mu}) = \frac{1}{\sqrt{\frac{2\pi}{N}}\,\sigma} \exp\left[-\frac{1}{2}\frac{(\hat{\mu}-\mu)^2}{\frac{\sigma^2}{N}}\right] \tag{3-51}$$

The estimate error $e = \hat{\mu}-\mu$ has probability density function

$$p(e) = \frac{1}{\sqrt{\frac{2\pi}{N}}\,\sigma} \exp\left[-\frac{1}{2}\frac{e^2}{\frac{\sigma^2}{N}}\right] \tag{3-52}$$

By the definition of probabilities presented in Section II, the probability that the absolute error exceeds a specified bound e_d is precisely calculable from the error probability density function equation (3-52). Explicitly,

$$P\left(\{|e|>e_d\}\right) = \int_{-\infty}^{-e_d} p(e)de + \int_{e_d}^{\infty} p(e)\,de$$

$$= 2\int_{e_d}^{\infty} p(e)\,de \tag{3-53}$$

Table 4 gives evaluations of this integral for selected values of e_d normalized

Table 4

CONFIDENCE PROBABILITY TABLE, $\sigma_e = \sigma N^{-1/2}$

e_d/σ_e	$P(\{\|e\| > e_d\})$	$P(\{\|e\| < e_d\})$
.5	.616	.384
1	.318	.683
2	.046	.954

by $\sigma_e=\sigma N^{-1/2}$. Given $e_d = 2\sigma_e$, for example, we say that there is only a 4.6% probability that our estimate error magnitude exceeds $2\sigma_e$. Equivalently, we say that we have 95.4% confidence that our estimate error magnitude is less than $2\sigma_e$.

An important aspect of the preceding analysis if that two quantities must be enumerated to indicate the quality of the estimate. The required quantities are (1) the error bound and (2) the probability that the specified bound is (or, is not) exceeded. Generally one of several alternative approaches is used. Calculation of the confidence from an a priori error bound is one approach commonly used. Otherwise, a particular confidence is specified and the corresponding error bound is calculated. An approach we shall use entails specifying both the error bound and the confidence. Once these quantities are specified, equations (3-52) and (3-53) are used to calculate a minimum number of observations N that must be made to obtain the desired quality estimate.

An identical analysis procedure is applicable to the calculation of confidence bounds for power spectral densities and cross spectral densities. Results presented in references 2, 8, and 12 are summarized.

The underlying probability of measured signals of dynamic systems tends to be a normal distribution as demonstrated by the central-limit theorem. It follows that the real and imaginary parts of a signal DFT at each component frequency are independent, normally distributed random variables with zero means and equal variances (reference 2, p. 189). The PSD is calculated from the real and imaginary DFT components by equation (3-39) or

$$\Phi_{xx}\left(\frac{k}{T}\right) = \frac{2CF}{N^2}\,|X_p(k)|^2\,T$$

$$= \frac{2CF}{N^2}\,T\left\{R_e{}^2\,[X_p(k)] + Im^2\,[X_p(k)]\right\} \qquad (3\text{-}54)$$

As the sum of squares of normal random variables, $\Phi_{xx}\left(\frac{k}{T}\right)$ is a chi-square random variable with two degrees of freedom (reference 4, p. 250). The ratio of the standard deviation to the mean of a chi-square distribution having n degrees of freedom is a constant given by

$$\frac{\sigma}{\mu} = \sqrt{\frac{2}{n}} \tag{3-55}$$

Chi-square probability density functions for higher degree of freedom are plotted in figure 25.

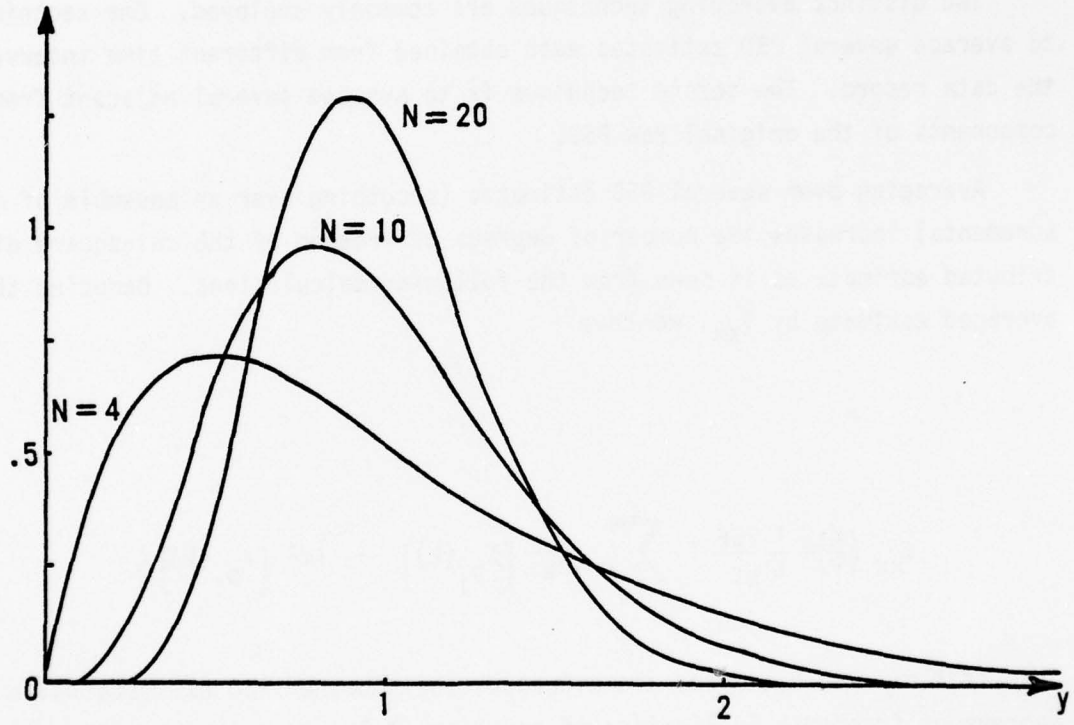


Figure 25. Chi-Square Probability Density Functions for Several Degrees of Freedom.

Clearly, our PSD estimate is a chi-square random variable with two degrees of freedom. As a consequence of equation (3-55) with n equal to 2, the PSD estimate given by equation (3-54) has a variance equal in magnitude to the estimate itself. This estimation error is intolerably large for most applications.

*Roughly speaking, the number of degrees of freedom is the number of squares of normal random variables summed to form the chi-square variable. That is $x^2 = x_1^2 + x_2^2 + \ldots x_n^2$ has n degrees of freedom. See Papoulis (reference 4, p. 250).

AD-A038 356 AIR FORCE WEAPONS LAB KIRTLAND AFB N MEX F/G 12/1
DATA ANALYSIS FOR STOCHASTIC PROCESSES.(U)
FEB 77 J E NEGRO
UNCLASSIFIED AFWL-TR-76-193 NL

2 OF 2
AD
A038356

END
DATE
FILMED
5-77

The estimation errors of the raw PSDs computed by equation (3-54) are reduced by averaging techniques. Averaging, in effect, increases the number of degrees of freedom of the chi-square random variable or PSD. By equation (3-55) this averaging process reduces the estimated variance in proportion to the estimate mean, a constant. The variance reduction is proportional to $\ell^{-1/2}$ where ℓ is the number of independent raw PSDs averaged to obtain the estimate.

Two distinct averaging techniques are commonly employed. One technique is to average several PSD estimates each obtained from different time intervals of the data record. The second technique is to average several adjacent frequency components of the original raw PSD.

Averaging over several PSD estimates (smoothing over an ensemble of measurements) increases the number of degrees of freedom of the chi-square distributed estimate as is seen from the following calculations. Denoting the averaged estimate by $\hat{\Phi}_{xx}$, we have

$$\hat{\Phi}_{xx}\left(\frac{K}{T}\right) = \frac{1}{\ell}\frac{2CF}{N^2}T\sum_{i=1}^{\ell}\left\{Re^2\left[X_{p_i}(k)\right] + Im^2\left[X_{p_i}(k)\right]\right\} \tag{3-56}$$

Then if ℓ raw PSD estimates are averaged, the smoothed PSD estimate has $n = 2\ell$ degrees of freedom. Application of equation (3-55) to this average shows that the standard deviation of the smoothed estimate is $\ell^{-1/2}$ times less than that of the original estimate.

Similar results are obtained for smoothed estimates derived as averages of adjacent frequency components of a single raw PSD. This averaging technique is particularly advantageous whenever a Φ_{xx} (or log Φ_{xx}) versus log-frequency presentation of the PSD is employed as is commonly the case (reference 8, pp. 90-93). The spacing of PSD estimates uniformly in frequency as occurs with DFT computations gives a far greater resolution that can be legibily displayed on a logarithmetic frequency graph. These 'extra' PSD estimates can be averaged to reduce the uncertainty of the estimate. For frequency averaging,

$$\Phi_{xx}\left(\frac{k+\ell/2}{T}\right) = \frac{1}{\ell}\frac{2CF}{N^2}T\sum_{i=1}^{\ell}\left\{R_e^2\,[X_p(k+i)] + \ell m^2\,[X_p(k+i)]\right\} \qquad (3\text{-}57)$$

An apparent disadvantage of the frequency smoothing technique in comparison with the ensemble smoothing technique of equation (3-56) is the loss of obtainable resolution, especially at low frequency ranges. However, the ensemble smoothing technique requires additional data samples. For an identical statistical confidence of PSD estimates given the same number of signal data points, each method yields the same frequency resolution. The ensemble smoothing technique is computationally less expensive, however. If N points are divided into ℓ intervals for which transforms and PSD's are computed and averaged, then the computational savings over frequency smoothing a N point sequence to obtain the same confidence is roughly a factor $\log_2 \ell$.

That the ensemble averaging and frequency smoothing yield the same confidence and frequency resolution given the same number of raw data points is illustrated by the following example.

EXAMPLE 3-1

Given $N = 4 \times 2^{10}$ data points, compute the frequency resolution of smoothed PSD estimates having a confidence equivalent to 8 degrees of freedom. A 1 KHz sampling frequency is used.

Method 1, Ensemble smoothing:

Four separate DFTs must be computed to yield the required degree of freedom. One fourth the total available samples, or 2^{10} points, is available for each DFT. The frequency resolution obtained is

$$\Delta f = \frac{1}{T} = \frac{1}{2^{10}/f_s} = 1.0 \text{ Hz}$$

Method 2, Frequency smoothing:

Four adjacent PSD frequency components must be averaged to yield an estimate with the required degrees of freedom. In this case, all N data points are available for the raw DFT computation. The unsmooth frequency resolution is

$$\Delta f' = \frac{1}{T} = \frac{1}{2^{12}/f_s} = .25 \text{ Hz}$$

This frequency resolution is 4 times smaller than available in Method 1 since the data interval used for the DFT calculation is 4 times longer. But we must average 4 adjacent frequency components to obtain the desired confidence.

$$\Delta f = 4 \cdot \Delta f' = 1.0 \text{ Hz}$$

Identical frequency resolution is obtained in either case.

Confidence bounds are computed from a normalized chi-square distribution. Normalizing this random variable by its mean we obtain multiplying factors which determine the upper and lower confidence bounds of PSDs as a multiple of the estimate itself. Small confidence bounds require that the multiplying factor be nearly equal to unity.

Upper and lower confidence bound multiplying factors tabulated in table 5

Table 5

CONFIDENCE INTERVAL UPPER AND LOWER BOUNDARY MULTIPLIERS, M, FOR AN n DEGREE OF FREEDOM CHI-SQUARE DISTRIBUTION

	100α:	80 Percent		90 Percent		95 Percent	
$2\ell=n$	M:	M_ℓ	M_u	M_ℓ	M_u	M_ℓ	M_u
2		.11	2.31	.05	3.00	.03	3.69
10		.49	1.60	.39	1.83	.33	2.05
20		.62	1.42	.54	1.57	.48	1.71
60		.77	1.24	.72	1.32	.68	1.39
120		.83	1.17	.80	1.22	.76	1.27

are derived from the normalized chi-square distribution. Both upper and lower factors must be specified since the chi-square density function is asymetric about its mean. Calculation of upper and lower multiplying factors for cases not presented in table 5 are calculable by methods illustrated in Example 3-2.

EXAMPLE 3-2

Given the percent confidence 100 α that we desire a PSD estimate to have and the number of degrees of freedom n equal to twice the number of raw DFT components averaged to calculate the PSD estimate, calculate the upper and lower confidence bound multipliers for the estimate. The estimate $\hat{\Phi}_{xx}$ has a chi-square distribution with mean Φ_{xx} and standard deviation [see equation (3-55)]

$$\sigma = \sqrt{\frac{2}{n}}\ \phi_{xx}$$

where n is the number of degrees of freedom and is equal to 2ℓ. Thus we need to calculate the upper and lower multiplier M_u and M_ℓ such that

$$P\left(\left\{\hat{\Phi}_{xx} > M_\mu\ \Phi_{xx}\right\}\right) + P\left(\left\{\hat{\Phi}_{xx} < M_\ell\ \Phi_{xx}\right\}\right) = 1-\alpha$$

To ensure an unbiased confidence interval we further require that

$$P\left(\left\{\hat{\Phi}_{xx} > M_\mu\ \Phi_{xx}\right\}\right) = P\left(\left\{\hat{\Phi}_{xx} < M_\ell\ \Phi_{xx}\right\}\right)$$

a new random variable $y = n\ \hat{\Phi}_{xx}/\Phi_{xx}$ is defined to facilitate later use of mathematical tables. Certainly, y has chi-square distribution with n degrees of freedom and mean n since it is simply a scalar multiple of $\hat{\Phi}_{xx}$. Now the multiples can be formulated as

$$P\left(\left\{\hat{\Phi}_{xx} > M_\mu\ \Phi_{xx}\right\}\right) = P\left(\left\{y > nM_\mu\right\}\right)$$

$$= \int_{nM_\mu}^{\infty} p(y)\ dy = \frac{1}{2}\ (1-\alpha)$$

where p(y) is the n degree of freedom chi-square probability density function.

Tabulation of nM_u given n and the integral value have been published (reference 13).

For the purpose of illustration, let $\alpha = .9$ corresponding to 90% confidence and suppose $n = 10$. Then from the tables we find that:

$$\int_{nM_u}^{\infty} p(y)\ dy = .05$$

for $nM_u = 18.3$ corresponding to $M_u = 1.83$. To compute the lower multiple M_ℓ using the tables, the problem must be recast into the form

$$1 - \int_{-\infty}^{nM_\ell} p(y)\, dy = \int_{nM_\ell}^{\infty} p(y)\, dy = 1 - \frac{1}{2}(1-\alpha)$$

$$\int_{nM_\ell}^{\infty} p(y)\, dy = .95$$

From the tables, $nM_\ell = 3.9$ or $M_\ell = 0.39$. These values agree with those tabulated in table 5. An identical calculation technique is applied when evaluating other cases.

Often PSD estimates are presented in a log magnitude versus log-frequency graphical format. For this presentation, a confidence interval 'spread' expressed in decibels is more convenient to use than the multipliers M_ℓ and M_u. Taking logarithms

$$S \triangleq 10 \log_{10} \frac{M_u}{M_\ell}$$

$$= 10 \log_{10}(M_u) - 10 \log_{10}(M_\ell) \qquad (3\text{-}58)$$

The spread S expressed in dB is tabulated in table 6 as a function of confidence

Table 6

CONFIDENCE INTERVAL SPREAD (dB) FOR AN n DEGREE OF FREEDOM CHI-SQUARE DISTRIBUTION

n	100α:	80 Percent	90 Percent	95 Percent
2		13.4	17.6	21.7
10		5.2	6.7	11.9
20		3.6	4.6	5.5
60		2.0	2.6	3.1
120		1.5	1.8	2.2

and the chi-square degrees of freedom of the estimate. Blackman (reference 14, p. 23) presents a fairly accurate formula for calculating the dB confidence spread associated with chi-square distributions. The formula, reasonably accurate for n > 10 is

$$S = \frac{K}{\sqrt{n-1}} \tag{3-59}$$

where K is determined by the desired confidence according to table 7. For a

Table 7

CONFIDENCE SPREAD CONSTANTS
(See equation (3-59))

100	K(dB)
80%	16
90%	20
95%	24
98%	29

given confidence, S is readily calculated as a function of n. Conversely, n can be expressed as a function of S; this technique is commonly employed to determine the number of terms that must be averaged to achieve a specified confidence and error tolerance bound.

Similar concepts may be applied toward determining confidence bounds for CSD's, coherence functions, and transfer function estimates. The procedures are significantly more difficult in these cases. The references present some of the details (reference 2, pp. 193-203). These details are not presented in this guide. Analyses for these cases show, as in the case of PSD estimates, that the quality of the estimate improves as more "raw" estimates are averaged to form the best estimate. Averages estimates for CSD's should be formulated in the same manner as averages for PSD's. Averages for coherence functions and transfer function estimates, however, should be computed from averaged PSD's and CSD's rather than by averaging raw coherence and transfer function estimates. This is readily seen to be valid in the case of coherence function estimates where the later procedure would always yield estimates identically equal to unity (see equation (3-43)), which is obviously an incorrect result. Furthermore, recall the technical development of section II.9 which demonstrated that transfer

functions estimates are only valid over the frequency ranges in which the coherence function tends toward unity. Otherwise, the "output" is not strongly influenced by the so-called "input" signal and a transfer function concept is inappropriate.

5. SPECIAL TOPICS

Several additional topics of importance in practical data analysis are discussed briefly in this section. The topics discussed are:

- Anti-aliasing filters
- Special considerations for CSD estimates
- Appending zeroes to signal sequences
- Bias and trend removal
- Sinusoidal signal components
- Transfer function estimates for deterministic inputs.

a. Anti-aliasing Filters

Anti-aliasing filters are required to prevent (or at least minimize) the aliasing or mimicking of low frequency signal components in a digitized signal by high frequency components in the original analog signal. The filtering must be accomplished prior to signal digitization. The filter attenuation beyond the Nyquist frequency (one-half the sampling frequency) must be sufficiently great that the amplitude of the attenuated high frequency component is small in comparison with the magnitude of the corresponding in-band frequency component so that aliasing distortion of a signal spectrum is negligible. This requirement dictates that multipole filters be used to give adequate out-of-band attenuation. Also, we require that these anti-aliasing filters have a minimum distortion effect on the signal spectrum for a reasonably large portion of the useable frequency band. For these reasons, "sharp cutoff" filters are used having break frequencies equal to approximately 40% of the sampling frequency f_s. Signal frequency components below $.4f_s$ are passed with negligible attenuation while components above $.4f_s$ are attenuated several orders of magnitude. Signal phase, on the other hand, is noticeably affected throughout the filter passband. This signal phase distortion is inconsequential for PSD estimates but becomes critical in CSD and transfer function estimates. In any event, the data analyst must understand the purpose of the anti-aliasing filter and the manner in which it affects the interpretation of data, especially if non-standard filters are chosen.

b. Special Considerations for CSD Estimates

Unlike PSD estimates, CSD estimates require calculations involving two distinct signals. The CSD estimate is sensitive to any process which affects the relative phase of the two signals. This phase distortion additively corrupts the CSD phase estimate. Special attention, then, must be paid to sources of relative phase distortion whenever the CSD phase information is important. One source of relative phase distortion is difference in signal phase due to use of different anti-aliasing filters having differing phase properties. Anti-aliasing filters with identical phase characteristics must be used if the CSD phase estimate is not to be distorted. Alternatively, the relative phase distortion between the filters can be measured or calculated and that result used to correct the CSD phase computation. Another primary source of relative phase distortion between two signals is time delay between the signals. A pure time delay has associated with it a phase delay that is linear in frequency. The phase difference is related to the time delay T_d by the equation:

$$\phi = -wT_d = -2\pi f\ T_d$$

Generally, this phase delay is intolerable at high frequencies for $T_d > (10f_s)^{-1}$. By equation (2-95), this phase component is induced into the CSD by the time delay transfer function. Sources of an equivalent time delay between two signals include inaccurate time bases for either or both signals, recording delays caused by recorder magnetic head skew, electronic recording and reproduce delays, and asynchronous sampling while digitizing. Whenever accurate phase information is required, these time delays must be minimized, and the residual delay must be measured so that it may be removed from the estimate a posteriori.

Another consequence of signal time shift is a reduction of apparent coherence in proportion to the ratio of time shift to the data interval time duration. As shown in reference 15:

$$\hat{\gamma}^2 = \gamma^2\ [1-\tau/T]^2$$

where τ is the time shift and T the data interval duration.

c. Appending Zeros to Signal Sequences

Several authors (references 2 and 12) recommend appending zeros to a data sequence to increase the total number of signal points to an integer of two. Such a total number of points is particularly suitable for the most common FFT algorithms which can operate only upon data sequences having 2^m points. Adding zeros effectively alters the intended window function and its desired spectral properties. These disadvantages outweigh any advantage gained by using additional data points. Alternatively, modern (reference 11, p. 307) FFT algorithms have been devised which effectively transform N point sequences for any N that is highly composite (i.e., N is the product of many factors). These methods should be employed in cases where maximum frequency resolution is required and hence the longest data interval possible must be transformed. Also, appending zeros to a data sequence will change the RMS level calculated for the signal interval.

d. Trend Removal

For many applications, the signal's mean value and the slope of a straight line approximation are important signal characterizations. The mean value identifies dc signal bias while the line slope identifies trends or a time correlation of the data. These values are calculated by the least-squares techniques presented in section II.5. Once these values are known, no additional information is gained by calculating the PSD of the data trend line component, as the corresponding PSD is uniquely parameterized by the dc and the trend slope. Trends must be removed if measurement of the low frequency random spectra is important to the analysis; otherwise, the low frequency spectra contributed by the trend can dominate the spectra contributed by the random signal component. Window functions can also be used to reduce the leakage distortion of linear trends (reference 8, p. 81). More generally, any deterministic signal component should be removed from the signal prior to further processing to characterize signal stochastic properties. Also, trend removal is important in preventing numerical inaccuracies that may arise when computing transforms on small word length computers.

Procedures for removing deterministic components from measured signals must be exercised judiciously so that only actual trends are removed. Bendat (reference 2, p. 291) recommends that the trends be removed only if the trend is "physically expected or clearly apparent in the data."

e. Sinusoidal Signal Components

Often, spectral techniques identify resonances or modes associated with the process being observed. Lightly damped resonances have corresponding PSD's exhibiting large spikes at one or more discrete frequencies. These spikes are associated with sinusoidal (or periodic) signal components. The resonant frequency and the energy associated with the sinusoid are typically the quantities of most interest to the data analyst. The resonant frequency is determined by the frequency at which the PSD has a local maximum. The uncertainty of the frequency estimate is the resolution bandwidth of the PSD, Δf equals ℓ/T for generalized frequency smoothed PSD estimates. The energy associated with the resonance is estimated by ascribing all the energy of the PSD spike to the resonance. This energy is obtained by the techniques developed in section II.9.

$$E(f_s) = \int_{f_s - \Delta f/2}^{f_s + \Delta f/2} \Phi_{xx}(f)df \qquad (3\text{-}60)$$

where f_s is the resonant frequency and Δf is approximately 5 times the transform fundamental frequency resolution. This bandwidth is required to include the spectral leakage side-lobe components as shown in figure 20b.

The amplitude of a sinusoid associated with the spectral resonance spike is estimated from the spike energy with the assumption that all the spike energy is associated with the energy of a sinusoid. A sinusoid with zero-to-peak amplitude A has energy $\frac{1}{2}A^2$. It follows that

$$A \approx \sqrt{2 \cdot E\,(f_s)} \qquad (3\text{-}61)$$

Note that the accuracy of the amplitude estimate is proportional to the square root of the energy estimate error. Thus a 10% energy error gives a 5% amplitude estimate error. Typically the energy estimate is high since it includes energy from the random signal component as well as the sinusoidal component. Good amplitude estimates are achieved provided that the PSD spike is an order of magnitude, or so, higher than the surrounding PSD floor, as shown in figure 26 for a typical spike.

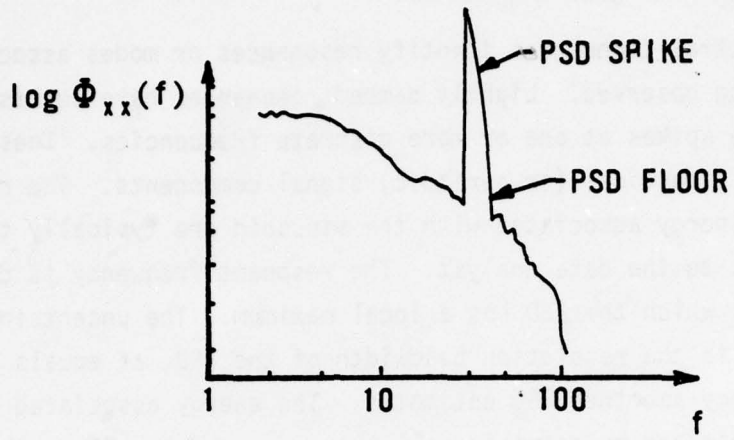


Figure 26. A Sinusoidal Signal PSD Spike.

Other methods based upon the absolute PSD spike amplitude are available for sinusoid amplitude estimation. These techniques are generally less accurate than the energy method because the spike amplitude depends upon the window function and the type of averaging employed. Additionally, the spike amplitude may vary as much as 50% depending upon the alignment of the true sinusoid frequency with the DFT frequencies.

f. Transfer Function Estimates for Deterministic Inputs

One has a natural tendency toward using the transform technique developed in this guide for determining system transfer functions given samples of deterministic system inputs and outputs. In view of the stochastic process results, we are further tempted to simply ratio output to input DFT's to obtain system transfer function estimates directly. These concepts can be applied provided that the analyst bears in mind the limitations of the technique and the possible gross estimate distortions that may result. These distortions arise from the aliasing and leakage distortion associated with DFT's. Transforms computed by this method correspond to systems having periodic steady state waveforms identical to the observed transient waveform. Problems associated with applying transform techniques to deterministic transfer function estimation are illustrated in the following example.

EXAMPLE 3-3

Consider the system in figure 27a excited by a step input initiated at time T_S.

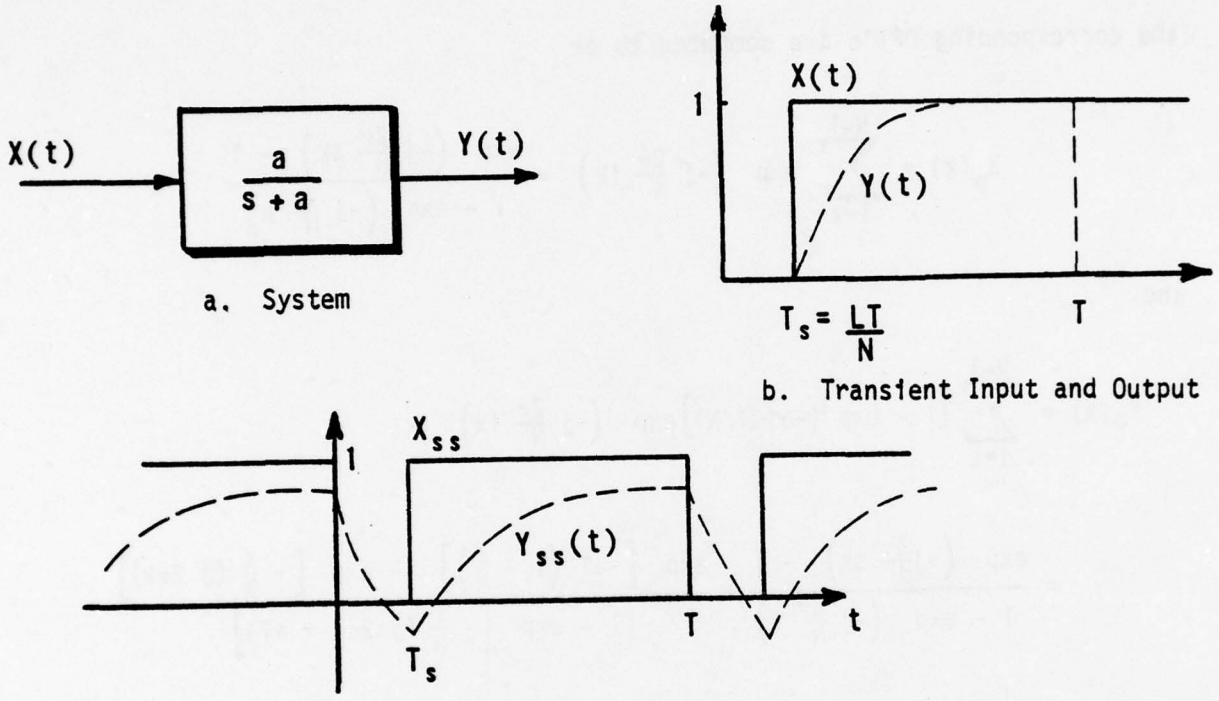


Figure 27. System Identification From Deterministic Input Output Data.

The process is observed from time zero until T with data samples taken T/N seconds apart. The output is

$$y(t) = \int_0^{\tau} e^{-a(t-\tau)} u_{-1}(\tau - T_s) d\tau$$

$$= \left[1 - e^{-a(t-T_s)}\right] u_{-1}(t-T_s)$$

where the function $u_{-1}(t-T_s)$ is defined

$$x(t) = u_{-1}(t-T_s) = \begin{cases} 1, & t \geq T_s \\ 0, & \text{otherwise} \end{cases}$$

the corresponding DFT's are computed to be

$$X_p(k) = \sum_{i=\ell}^{N-1} \exp\left(-j\frac{2\pi}{N} ik\right) = \frac{\exp\left(-j\frac{2\pi}{N}\ell k\right) - 1}{1 - \exp\left(-j\frac{2\pi}{N}k\right)}$$

and

$$Y_p(k) = \sum_{i=\ell}^{N-1} \left[1 - \exp\left(-ai\overline{T}/N\right)\right] \exp\left(-j\frac{2\pi}{N} ik\right)$$

$$= \frac{\exp\left(-j\frac{2\pi}{N}\ell k\right) - 1}{1 - \exp\left(-j\frac{2\pi}{N}k\right)} + \frac{\exp\left[-aT\left(1 - \frac{\ell}{N}\right)\right] - \exp\left[-\frac{\ell}{N}(j\,2\pi k)\right]}{1 - \exp\left[-\frac{1}{N}(j\,2\pi k + aT)\right]}$$

A transfer function estimate formed by ratioing $Y_p(k)$ and $X_p(k)$ is compared with the true transfer function $H(k) = \left(1 + j\frac{2\pi k}{aT}\right)^{-1}$ in figure 28. Note the peculariities that occur in even this most elementary example.

It follows from the Fourier series interpretation of the discrete Fourier transform that a perfect transfer function estimate is obtained only in the case that the transforms of steady state input and output waveforms as shown in figure 27c are used in the analysis, and that the input signal is bandlimited.

As illustrated in the example, particular care must be exercised when extracting function estimates from deterministic input output data. Similar care is not generally required with ergodic random data for which each analysis interval exhibits the random character of the entire data interval.

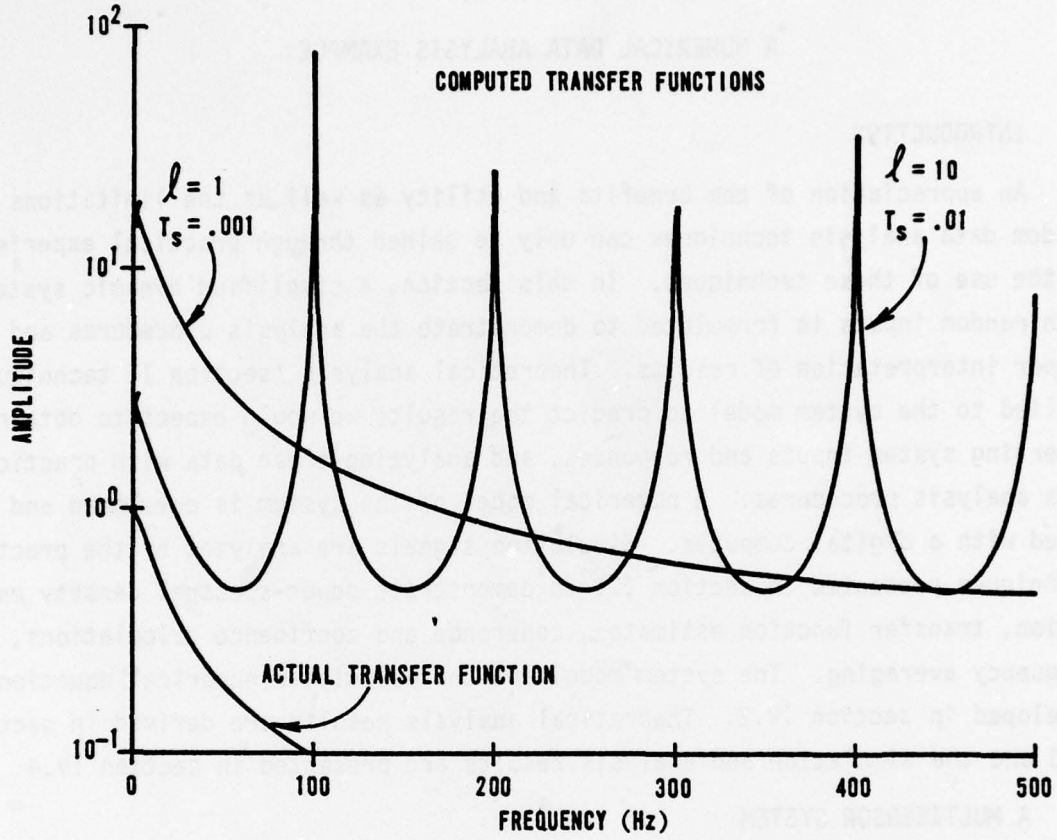


Figure 28. A Comparison of Transfer Function Estimates.

6. SUMMARY

Numerical algorithms required for analysis of digitized random data using a digital computer have been developed in this section. The digitization, the truncation of data sequences, and the statistical nature of the data all presented limitations in the scope in which our computational results could be interpreted. Frequency aliasing, leakage phenomena, and estimate confidence bands have been shown to have important consequence in data analysis.

SECTION IV

A NUMERICAL DATA ANALYSIS EXAMPLE

1 INTRODUCTION

An appreciation of the benefits and utility as well as the limitations of random data analysis techniques can only be gained through practical experience in the use of these techniques. In this section, a simplified dynamic system with random inputs is formulated to demonstrate the analysis procedures and the proper interpretation of results. Theoretical analysis (section II technique) is applied to the system model to predict the results we would expect to obtain by observing system inputs and responses, and analyzing these data with practical data analysis procedures. A numerical model of the system is developed and simulated with a digital computer. Simulation signals are analyzed by the practical techniques presented in section III to demonstrate power-spectral density estimation, transfer function estimates, coherence and confidence calculations, and frequency averaging. The system model and the associated numerical equations are developed in section IV.2. Theoretical analysis results are derived in section IV.3 and the simulation and analysis results are presented in section IV.4.

2 A MULTISENSOR SYSTEM

A suitable numerical example system must be sufficiently generalized to exemplify the significant aspects of the available random data analysis techniques. Simultaneously, the considered system must be simple enough to be amenable to direct theoretical analysis which could be compared to the simulated results. Finally we desire that the chosen example resemble the random data analysis situations which we encounter in practice.

Most generally in engineering practice, we are interested in characterizing random disturbances and the responses of dynamic systems to that disturbance. We are constrained to practical measurements of signals. System responses and disturbance phenomena are observed with imperfect sensors which contribute additional random components (sensor noise) to the observed or measured quantities. For example, the measurements and analyses of random, aerodynamically induced pressure fluctuations acting upon mechanical structures are directly affected by random signal components due to the pressure transducer, the instrumentation system, and the signal processing, as well as the desired signal directly attributable to actual pressure fluctuations.

Two distinct cases of sensor noise data corruption occur. In one case, a sensor is used for instrumentation purposes only. Sensor noise additively corrupts the measured quantity. Another case of importance occurs with sensor imbedded as a control sensor in a dynamic system. System dynamics as well as sensor noise must be known in these cases to properly account for the measured random phenomena. Examples of such sensors are pressure transducers of a hydraulic actuator or the tracking sensor of a pointing system.

The generalized system we choose to investigate includes one sensor of each type. A simple dynamic system described by a first-order differential equation is used. Sensor noise is modeled simply as white noise so as not to obscure important aspects of random data analyses with a multitude of parameters. Approximation of noise spectra by white noise is valid in many practical cases because the noise is "wide-band" in comparison with system dynamics. Effects of "colored" sensor noise are commented upon as appropriate in the theoretical derivations.

In addition to control loop sensor noise the dynamic system is stochastically excited by an exponentially correlated stochastic disturbance phenomena. Such a disturbance model is the simplest correlated (colored) disturbance we can encounter; yet it exemplifies the notions of correlatedness, bandwidth and non-constant spectra which we typically encounter, though often in a more complicated manner, in practice.

The complete system model used for our theoretical and simulation investigation is described by its transfer function block diagram given in figure 29. The control system output y is compared with the desired system output (reference input) r by the control system sensor. This sensor measures the control error accurately except for an additive sensor noise ν. The noise corrupted sensor output e_m is the only system error signal which can be instrumented and analyzed. The error signal is amplified by gain K; the resulting signal drives the system plant G. The system plant is also driven by the external disturbance d. In fact, the purpose of the control system is to maintain the plant output y at the desired level r (zero) despite perturbations induced by the disturbance.

The Laplace transform complex variable s is used in the representation of G and the stochastic disturbance model (see reference 6). The system plant is a simple integrator. The stochastic disturbance is modeled as the output of a single-pole low-pass filter. The filter has bandwidth $\lambda/2\pi$ (Hz) and unity dc gain. The disturbance model excitation and the sensor noises are modeled as <u>independent</u> normally distributed white noise stochastic processes.

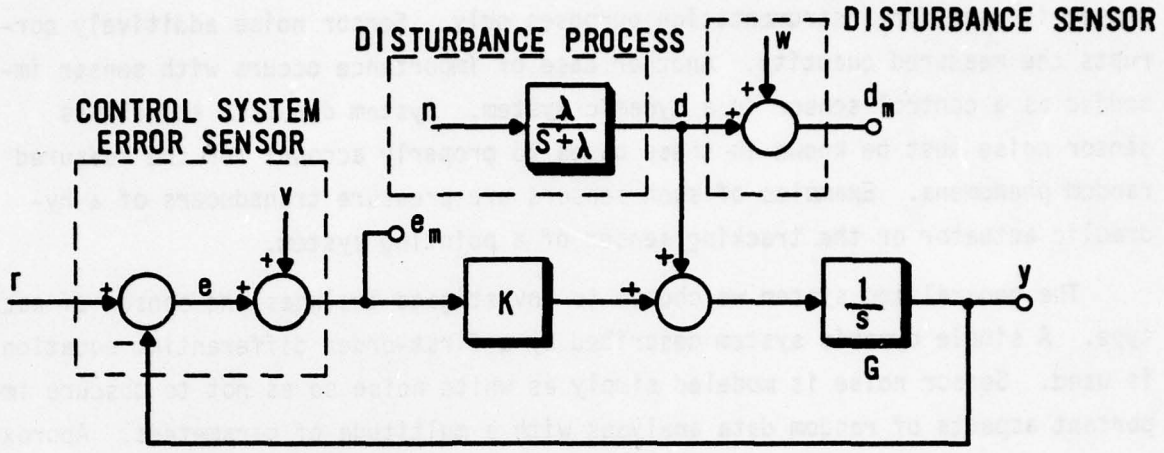


Figure 29. Stochastic Analysis Example Model.

Model Nomenclature

K - control system loop gain

r - dynamic system reference input; r(t) = o

y - control system output

e - actual system error

v - error sensor noise (white spectra, gaussian pdf)

e_m - measured system error

d - exponentially correlated disturbance phenomena

w - disturbance measurement sensor noise (white spectra, gaussian pdf)

d_m - measured system disturbance

n - disturbance model excitation (white spectra, gaussian pdf)

λ - disturbance model bandwidth parameter BW (in Hz) = $\lambda/2\pi$

In absence of disturbance phenomena, the control system of figure 29 has zero steady state error and output responses. Disturbances excite the system to give non-zero error and system output. RMS characterizations of these quantities are a useful figure of merit indicating the performance of the system. The smaller the RMS of these quantities, the better the system performance.

Equations modeling the analysis system are presented in the remainder of this section. In section IV.3 theoretical stochastic relationships among the various signals are developed. Finally, in section IV.4 numerical results are presented, interpreted, and compared with the theoretically expected results.

The disturbance process is modeled by a first order differential equation.

$$\dot{d} = \lambda(n-d) \tag{4-1}$$

The true system disturbance is represented by d. The constant λ is a disturbance bandwidth parameter and n is a white noise excitation of the model. Only a noise corrupted signal is available for measurement. The noisy disturbance measurement is modeled by simply adding a white noise signal component w to the true system disturbance. Denoting the disturbance measurement by d_m gives the following algebraic relationship.

$$d_m = d + w \tag{4-2}$$

The control system is modeled by the following signal relationships evident from figure 29

$$\dot{y} = Ke_m + d = K(v-y) + d \tag{4-3}$$

and the algebraic relationships

$$e_m = e + v \tag{4-4}$$

and

$$e = r-y = -y \qquad (r=o) \tag{4-5}$$

Differential equations (4-1) and (4-3) and the algebraic equations (4-2), (4-4), and (4-5) specify the structure of the system. The model is completely characterized by additionally specifying numerical values for the parameters λ and K and by specifying the statistics of the white noise process n, w and v. These white noise processes are chosen to have zero means and spectral densities 2, Φ_{ww}, and

Φ_{vv}, respectively. All combinations of each signal having either a low or a high spectral density make up the four cases considered. Parameters λ and K are given values $2\pi 50$ and $2\pi 10$ respectively.

The equations presented here are not suitable for numerical simulation. The differential equations must be approximated by finite-difference equations. These equations are obtained by approximating derivatives with a finite difference.

$$\dot{x} \approx \frac{1}{\Delta T}\,[x(t) - x(t-\Delta t)] \tag{4-6}$$

Substituting this approximation into differential equations (4-1) and (4-3) leads to the finite difference equations

$$d_k = d_{k-1} + \Delta T\ \lambda(n_{k-1} - d_{k-1}) \tag{4-7}$$

and

$$y_k = y_{k-1} + \Delta T\ K(u_{k-1} - y_{k-1}) + \Delta T\ d_{k-1} \tag{4-8}$$

The subscript k denotes the successive increments of the discrete process and relates to time $t = k\ \Delta T$ in the continuous process.

$$d_k \approx d(k\Delta T) \tag{4-9}$$

The discrete data algebraic equations are simply:

$$(d_m)_k = d_k + w_k \tag{4-10}$$

$$(e_m)_k = e_k + u_k \tag{4-11}$$

and

$$e_k = -y_k \tag{4-12}$$

Finally, the 'white' noise sequences n_k, v_k, and w_k are obtained from a random number generator. Samples from the generator have variances $Q_n/\Delta T$, $Q_v/\Delta T$, and $Q_w/\Delta T$ respectively and are independent of each other and from sample to sample. Normalization of the variances by ΔT is required for an accurate approximation to 'continuous' white noise which has infinite variance. The random number generator described in reference 16 is used in the simulation. It produces independent samples having normal distributions. Three separate starting seeds are used to produce the three independent sequences.

The finite-difference equations used to produce signal sequences for stochastic signal analysis are summarized in table 8. These equations approximate the

Table 8.

SUMMARY EQUATIONS FOR THE DISCRETE TIME STOCHASTIC SYSTEM NUMERICAL EXAMPLE

DISTURBANCE PROCESS MODEL

$$d_k = d_{k-1} + \Delta T \cdot \lambda \cdot (n_{k-1} - d_{k-1})$$

CONTROL SYSTEM MODEL

$$y_k = y_{k-1} + \Delta T \cdot K(v_{k-1} - y_{k-1}) + \Delta T\, d_{k-1}$$

SYSTEM ERROR SENSOR MODEL

$$(e_m)_k = e_k + v_k = - y_k + v_k$$

DISTURBANCE SENSOR MODEL

$$(d_m)_k = d_k + w_k$$

White noise sequences n_k, v_k, and w_k generated by a gaussian number generator to have variances $Q_n/\Delta T$, $Q_v/\Delta T$, and $Q_w/\Delta T$, respectively.

differential equations describing a single-pole control system excited by feedback sensor noise and a stochastic disturbance. The approximation is very good

at low frequencies; it suffers from aliasing like distortions at higher frequencies. The discrete time model has greater phase lag and less magnitude attenuation at frequencies approaching one half the discrete equation update frequency. The equations given in Table 8 represent a complete mathematical description of the process under investigation. In section IV.3, the theoretical analyses of section II are applied to determine the PSD's, transfer function, and coherence functions we expected to calculate from the data. These theoretical analyses are possible in this case since we have perfect knowledge of the system and the statistics of the component stochastic signals.

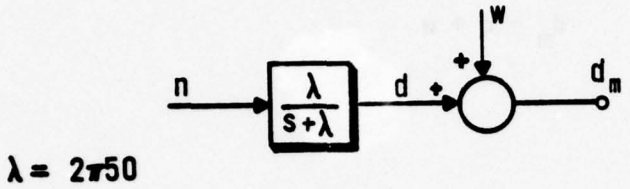
3 THEORETICAL COMPUTATIONS

Given the mathematical system model characterized by differential equations (4-1) and (4-3) along with algebraic relationships equations (4-2), (4-4), and (4-5), and applying to linear system relationships summarized in table 1 and equation (2-8) we can make theoretical calculations of signal PSD and coherence functions we should expect to calculate by random data analysis procedures. These theoretical results are compared with the signal analysis results as the latter are developed in section IV.4.

Two aspects of the stochastic analysis model example are investigated. First, the disturbance process model is studied; PSD's of the disturbance sensor output signal d_m are computed for several signal-to-noise ratios obtained by varying the amplitude of the wideband sensor noise signal w. Transfer function estimates for the disturbance process d/n are estimated from the quantity n (not measurable in practice) and the measured disturbance d_m. Transfer function methods for several RMS levels of w are compared. Secondly, transfer function estimates as well as PSD and coherence calculations are made from the measurement of system error e_m and the measured system disturbance d_m. These quantities are also computed for a variety of cases in which the signal to noise ratio of each sensor is varied over a 100dB range.

The basic analysis tool is the transfer function $H(2\pi f)$. Two sybsystems of the numeric example are analyzed. We first consider the simple process relating measured disturbance d_m to the white noise excitation n. We realize that a similar analysis is not possible with a practical system since no signal corresponding to n is physically available. However, we proceed with analysis of this simple process as it shall lend insight to the use and interpretation of our procedures. Next we shall consider the process relating system measure error e_m with

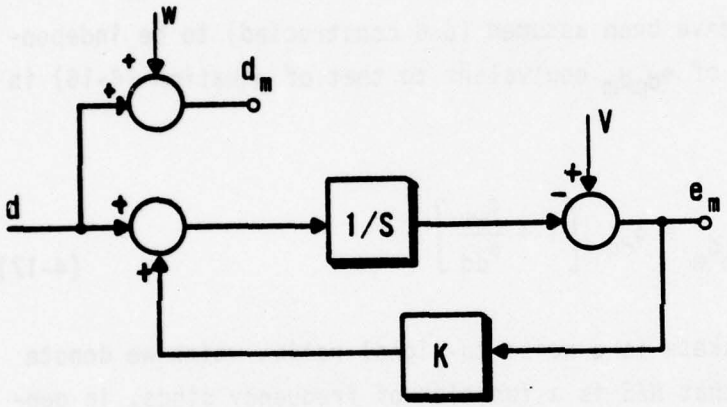
the measured disturbance d_m. These functional relationships are depicted explicitly in block diagrams of figure 30.



$\lambda = 2\pi 50$

Signals Available for Analysis
n
d
d_m

a. Disturbance Process Model



$K = 2\pi 10$

Signals Available for Analysis
d_m
e_m

b. Control System Model

Figure 30. Theoretical Process Transfer Functions.

The transfer function from input n to output d is seen to be

$$H(w) = \frac{\lambda}{\lambda + jw} \tag{4-13}$$

It follows from equation (2-80) that

$$\Phi_{dd}(f) = |H(2\pi f)|^2 \Phi_{nn}(f)$$

$$= \left[1 + \left(\frac{2\pi f}{\lambda}\right)^2\right]^{-1} \Phi_{nn}(f) \tag{4-14}$$

The additive noise w adds spectral component Φ_{ww} to the measured disturbance signal. Hence, the relationship

$$d_m = d + w \tag{4-15}$$

implies

$$\Phi_{d_m d_m} = \Phi_{dd} + \Phi_{ww} \tag{4-16}$$

because the processes n and w have been assumed (and constructed) to be independent. A useful representation of $\Phi_{d_m d_m}$ equivalent to that of equation (4-16) is obtained by factoring Φ_{dd}.

$$\Phi_{d_m d_m} = \Phi_{dd} \left[1 + \frac{\Phi_{ww}}{\Phi_{dd}} \right] \tag{4-17}$$

The second term within the brackets is a noise-to-signal ratio, which we denote notationally as N/S. Observe that N/S is a function of frequency since, in general, both Φ_{ww} and Φ_{dd} are functions of f. Over those frequency bands in which N/S << 1, the measurement spectrum closely approximates the spectrum of the desired signal d. The noise is inconsequential in these cases. Over the remaining frequency bands, the noise term contributes significantly to the measured spectrum. Substituting equation (4-14) into equation (4-17) gives the result

$$\Phi_{d_m d_m}(f) = \Phi_{nn} |H(2\pi f)|^2 \left[1 + \frac{\Phi_{ww}}{|H(2\pi f)|^2 \, \Phi_{nn}} \right] \tag{4-18}$$

This PSD is plotted in figure 31 for several ratios of Φ_{ww}/Φ_{nn}. In the examples considered, Φ_{nn} and Φ_{ww}, are constants because they represent spectral densities of 'white' noise processes. Thus, a low signal-to-noise condition, as previously defined, exists whenever

$$|H(2\pi f)|^2 > \frac{\Phi_{ww}}{\Phi_{nn}} \tag{4-19}$$

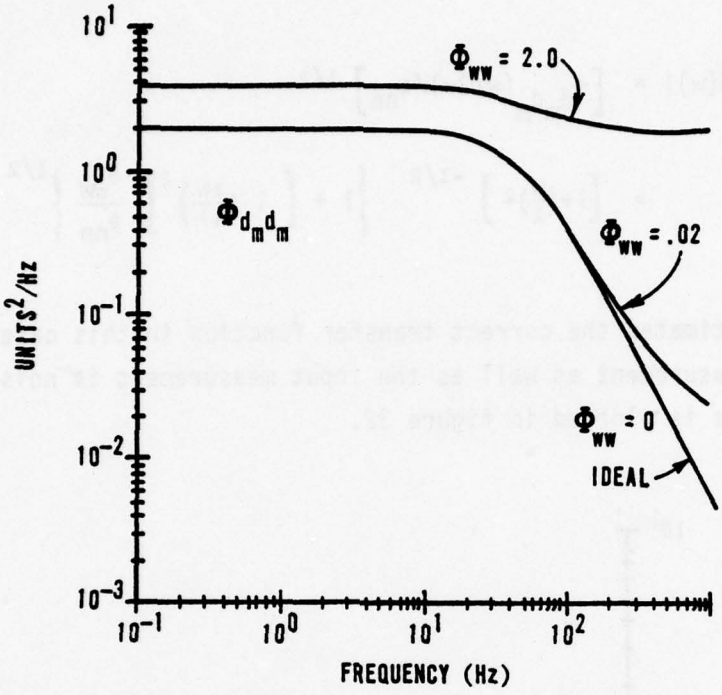


Figure 31. Disturbance Measurement Power Spectral Densities.

In addition to PSD's, we are also interested in transfer function estimates and the coherence function. These functions are computed by use of the equations developed in section II.9. The measured disturbance signal is taken as the transfer function output. The input is the white noise excitation n (see figure 30a).

TRANSFER FUNCTIONS

METHOD 1

$$\hat{H}(w) = \Phi_{d_m n}\ (w/2\pi)/\Phi_{nn}$$

$$= \frac{\lambda}{\lambda+jw} \qquad (4\text{-}20)$$

Method 1 yields an ideal estimate. The output noise, as long as it is incoherent with the system excitation, has no effect on the transfer function estimate.

METHOD 2

$$|\hat{H}(w)| = \left[\Phi_{d_m d_m}(w/2\pi)/\Phi_{nn}\right]^{1/2}$$

$$= \left[1+(\frac{w}{\lambda})^2\right]^{-1/2} \left\{1 + \left[1 + (\frac{w}{\lambda})^2\right] \frac{\Phi_{ww}}{\Phi_{nn}}\right\}^{1/2} \qquad (4\text{-}21)$$

Thus Method 2 estimates the correct transfer function in this case only when the output signal measurement as well as the input measurement is noise free. The Method 2 estimate is plotted in figure 32.

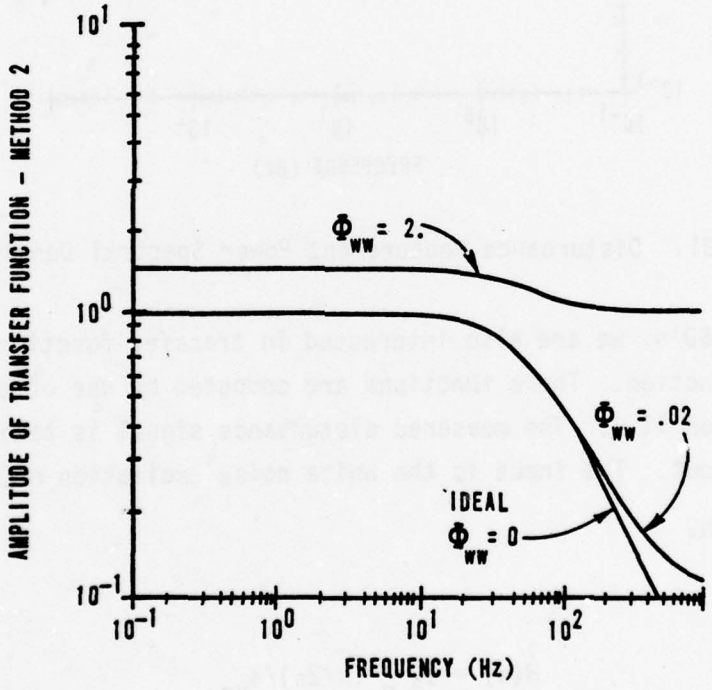


Figure 32. Disturbance Process, Method 2 Transfer Function Estimates.

COHERENCE FUNCTION

The coherence function $\gamma^2(f)$ is calculated by using equation (2-109)

$$\gamma^2_{d_m n}(f) = \frac{\left|\Phi_{d_m n}(f)\right|^2}{\Phi_{d_m d_m}(f)\ \Phi_{nn}}$$

$$= \left\{1 + \left[1 + \left(\frac{2\pi f}{\lambda}\right)^2\right] \frac{\Phi_{ww}}{\Phi_{nn}}\right\}^{-1}$$

$$\lambda = 2\pi 50 \tag{4-22}$$

We see that if $\Phi_{ww} << \Phi_{nn}$ that γ^2 is approximately unity at frequencies less than 50 Hz. At higher frequencies, γ^2 becomes smaller. The theoretical coherence function is plotted in figures 39 and 40 of section IV.4 in comparison with a coherence function calculated from the actual data. Observe that the coherence function is near unity over those frequency bands in which both transfer function estimates are accurate. Also, in the same frequency bands the coherent output power dominates the sensor noise.

Identical procedures are employed to calculate the power spectral density $\Phi_{e_m\ e_m}$. In this linear system, the error e_m is made up of two independent components - one due to the disturbance d and the other due to the sensor noise v. The appropriate transfer functions are

$$\frac{e_m}{d} \triangleq H_1(w) = \frac{1/K}{1 + jw/K} \tag{4-23}$$

and

$$\frac{e_m}{v} \triangleq H_2(w) = \frac{jw/K}{1 + jw/K} = jwH_1\ (w) \tag{4-24}$$

By superposition and from the linear system theorem we have

$$\Phi_{e_m e_m} = |H_1(2\pi f)|^2\ \Phi_{dd} + |H_2(2\pi f)|^2\ \Phi_{vv} \tag{4-25}$$

Notice that in the latter case, the precise manner by which the error sensor noise corrupts the error measurement is influenced by the control system dynamics.

This is a general result true for any sensor embedded as a feedback sensor in a control system. Plots of the theoretical $\Phi_{e_m e_m}$ are presented in figure 33.

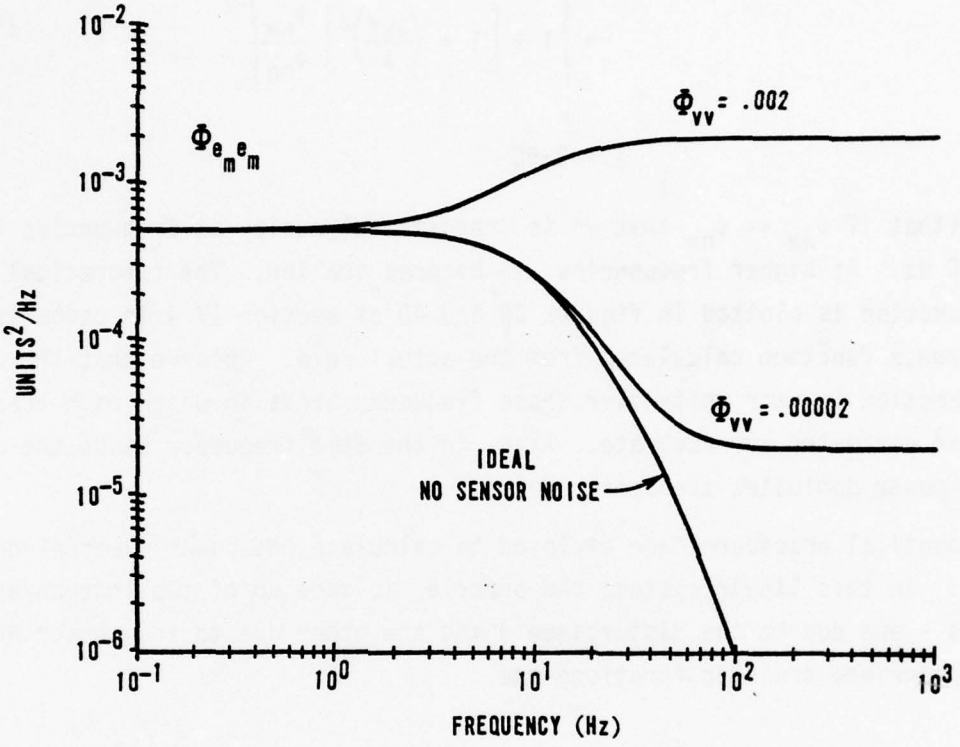


Figure 33. Control Error Measurement PSD's.

The measured signals d_m and e_m may be considered as the input and output, respectively, of a linear system. These signals, each corrupted by additive noise components, represent the general signal analysis transfer function estimation problem encountered in practice. *Theoretical transfer function estimates* based on the model transfer function and the noise characterizations are

IDEAL

$$H(w) = \frac{1}{K} \frac{1}{1 + jw/K} \qquad (4\text{-}26)$$

METHOD 1

$$\hat{H}(w) = \frac{\Phi_{e_m d_m}}{\Phi_{d_m d_m}} = \frac{H(w)\Phi_{dd}}{\Phi_{d_m d_m}} \tag{4-27}$$

$$= H(w) \left[1 + \frac{\Phi_{ww}}{\Phi_{dd}}\right]^{-1} \tag{4-28}$$

The last step follows from equation (4-17). We again note the importance of a noise-to-signal ratio, N/S. In this case the noise is characterized by Φ_{ww} while Φ_{dd} characterizes the signal. Over low N/S frequency bands the transfer function estimate is accurate. Conversely, over high N/S frequency bands the estimate is poor. Observe that the estimated transfer function magnitude is less than the true magnitude. Also note that the estimated transfer function is a real scalar multiple of the actual transfer function. Consequently, the phase component of the estimate remains accurate despite the noise-to-signal level.

METHOD 2

$$|\hat{H}(w)| = \left| \frac{\Phi_{e_m e_m}}{\Phi_{d_m d_m}} \right|^{1/2}$$

$$= H(w) \left[\frac{1+w^2 \dfrac{\Phi_{vv}}{\Phi_{dd}}}{1 + \dfrac{\Phi_{ww}}{\Phi_{dd}}} \right]^{1/2} \tag{4-29}$$

Only the transfer function magnitude can be estimated by Method 2. As for the Method 1 estimate, this estimate magnitude differs from the true transfer function magnitude by a scalar multiple. In this case, the multiple may be either less or greater than unity. The numerator terms represent an output signal noise-to-signal ratio. The w^2 factor appears in this term to compensate for the fact that the signals characterized by Φ_{vv} and Φ_{dd} occur in different locations of the control system; to properly compare them, they must be transferred to a

common comparison point. The output (see figure 30a) is chosen as the common comparison point. Observe that the relative transfer function from d or v to the output e_m is simply 1/s corresponding to $|H|^2 = 1/w^2$. This factor is the appropriate multiplier of Φ_{dd} so that it can be compared, on a noise-to-signal basis, with Φ_w. Similarly, the denominator factor also has a N/S term. An accurate transfer function estimate is obtained whenever the N/S ratios are much less than unity. For larger N/S ratios, the estimated transfer function deviates from the actual transfer function (except for the pathological case wherein the numerator and denominator N/S ratios vary so as to maintain a unity multiplying error factor).

COHERENCE FUNCTION

$$\gamma^2_{d_m e_m}(f) = \frac{|\Phi_{d_m e_m}(f)|^2}{\Phi_{d_m d_m} \Phi_{e_m e_m}}$$

$$= \left\{\left[1 + \frac{w^2 \Phi_{vv}}{\Phi_{dd}}\right] \left[1 + \frac{\Phi_{ww}}{\Phi_{dd}}\right]\right\}^{-1} \qquad (4\text{-}30)$$

Observe that the coherence function is made up of N/S factor terms. As either transfer function estimate (Method 1 or 2) deviates from the actual transfer function, γ^2 reduces in value from unity. Also note that γ is simply the ratio of the Method 1 to the Method 2 transfer function amplitudes.

The transfer function estimates and the coherence function theoretical values are plotted in comparison with transfer functions and coherence functions calculated from the simulated model data in section IV.4. We note the general result that typically, as the system noise increases, these functions diverge from their desired values.

4. SIMULATION AND ANALYSIS RESULTS

Results of applying the numerical analysis procedures developed in sections II and III to the simulated control system using the SIGANAL software are presented. These results are compared with theoretical values to demonstrate representative accuracies and interpretative procedures useful in analyzing PSD and transfer function estimate data.

For the hypotehtical system used in this simulation, two comparisons of the results are possible and equally instructive. The analysis results are compared to ideal values. Transfer function estimates, for example, are compared to the actual system transfer functions. Also, analysis results are compared to those theoretically expected results as derived in section IV.3. This later comparison is only possible in this simulated example, since the theoretical calculations require precise knowledge of system transfer functions and measurement noise properties which are not known in practice.

The system was simulated by iterating the modeling equations in table 8. The numeric values obtained for the disturbance process n_k, d_k, $(d_m)_k$, w_k, and the control system v_k, y_k and $(e_m)_k$ were written in blocks on an engineering unit (EU) format data tape for subsequent analysis by the SIGANAL software (reference 1).

The simulated signals were obtained with a finite difference update rate of 1 k Hz corresponding to a simulation of time increment ΔT = .001 second. The equations were iterated to give 8192 signal values simulating T = 8.2 seconds of system response.

The simulated signals stored on EU tape are analyzed by the SIGANAL software. The SIGANAL summary list and NAMELIST input cards used for the analysis are listed in table 9. Excerpts of these analysis results are presented and discussed in the remainder of this section.

a. Disturbance Process PSD's

The results of the disturbance process analysis presented in figures 34-40 are first reviewed. The disturbance process excitation signal measured PSD is plotted in figures 34a and b. These measured values differ somewhat from the actual value Φ_{nn} = 2.0 also plotted because of the statistical uncertainty of PSD estimates made from a finite duration observation of the signal as discussed in section III.4. Confidence bounds for 95% confidence are also plotted and were derived by use of the methods illustrated in Example 3-2. The PSD estimate in figure 34a is obtained from a minimum of eight frequency averages while the PSD in figure 34b, which varies more from the theoretical value, is obtained with a minimum of only four averages. The 95% confidence bound calculation results are summarized in table 10 for each of these cases. Case A employs at least four times the number of raw PSD estimates per average than Case B. The spread of the SIGANAL PSD estimates scale with the confidence bound calculations. Comparisons

Table 9

REQUIRED SIGANAL (VERSION 4) DATA CARDS FOR SIMULATION SIGNAL PROCESSING

```
      READFU(X,Y,ID),ANAL(Y,T,M,CH,RM,CF,SD,CSD,INST),
      PDF(Y,MIDP,PDIST,FO,ERPF,CHI),
      TAPDC(Y,O),FFTFCS(Y,F,R,I),CSD(R,I,SR,SI,FO,AM,PH,COH,LAB),
      PSD(R,T,FO,PW,CPW),TRFNT(R,T,SR1,SI1,FO,TRA,TRP,LAB),
      TFNT(PW,AM,PH,TA,TP,FO,SPW,LAB),
      PLOT(FO,AM,LAB),PLOT(FO,PH,LAB),PLOT(FO,COH,LAB),
      PLOT(FO,PW,ID),PLOT(FO,TRA,LAB),PLOT(FO,TRP,LAB),PLOT(MIDP,PD,TD),
      PLOT(FO,TA,LAB).
$FILES IDATA=1 $
$DATA INCR=-1,METLF(1)=2,2,-1,IBUG=1,MPTS=8192,
 PW(1)=1,4,10,25, CHANGE(1)=20,100,250, IFC=500,
 NTYPE(1)=4,3,3,4,4,4,1,4, $END
$FILES IDATA=1, $
$DATA METLF(1)=2,7,-1,7,6,-1,9,14,-1,
  14,13,-1,16,21,-1,21,26,-1,23,28,-1,
  28,27,-1,1, $END
$FILES IDATA=1, $
$DATA METLF(1)=3,2,-1,47*1,PW(1)=.25,.5,2.5,
 CHANGE(1)=20,100, IFC=250,$
$FILES IFTLF=-999, $
```

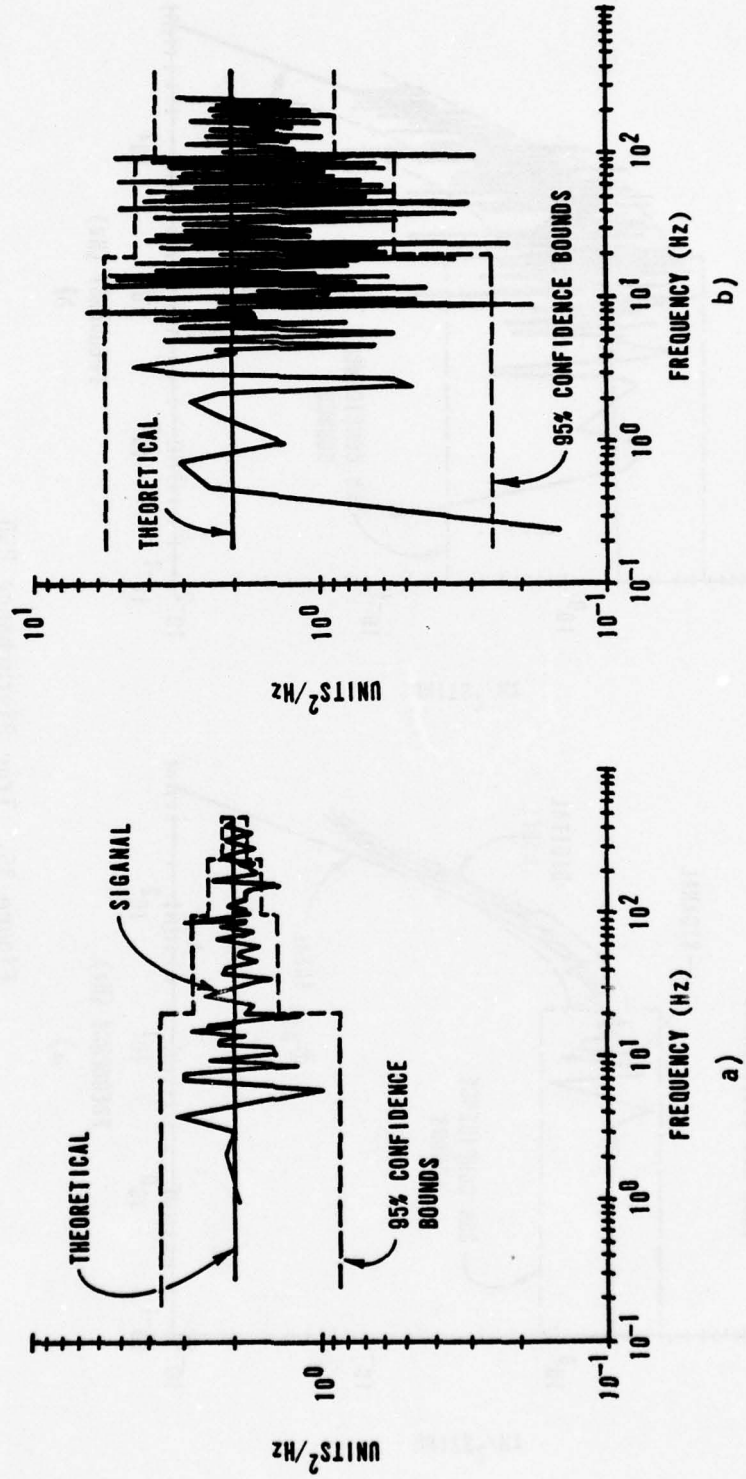


Figure 34. Disturbance Process Excitation PSD's.

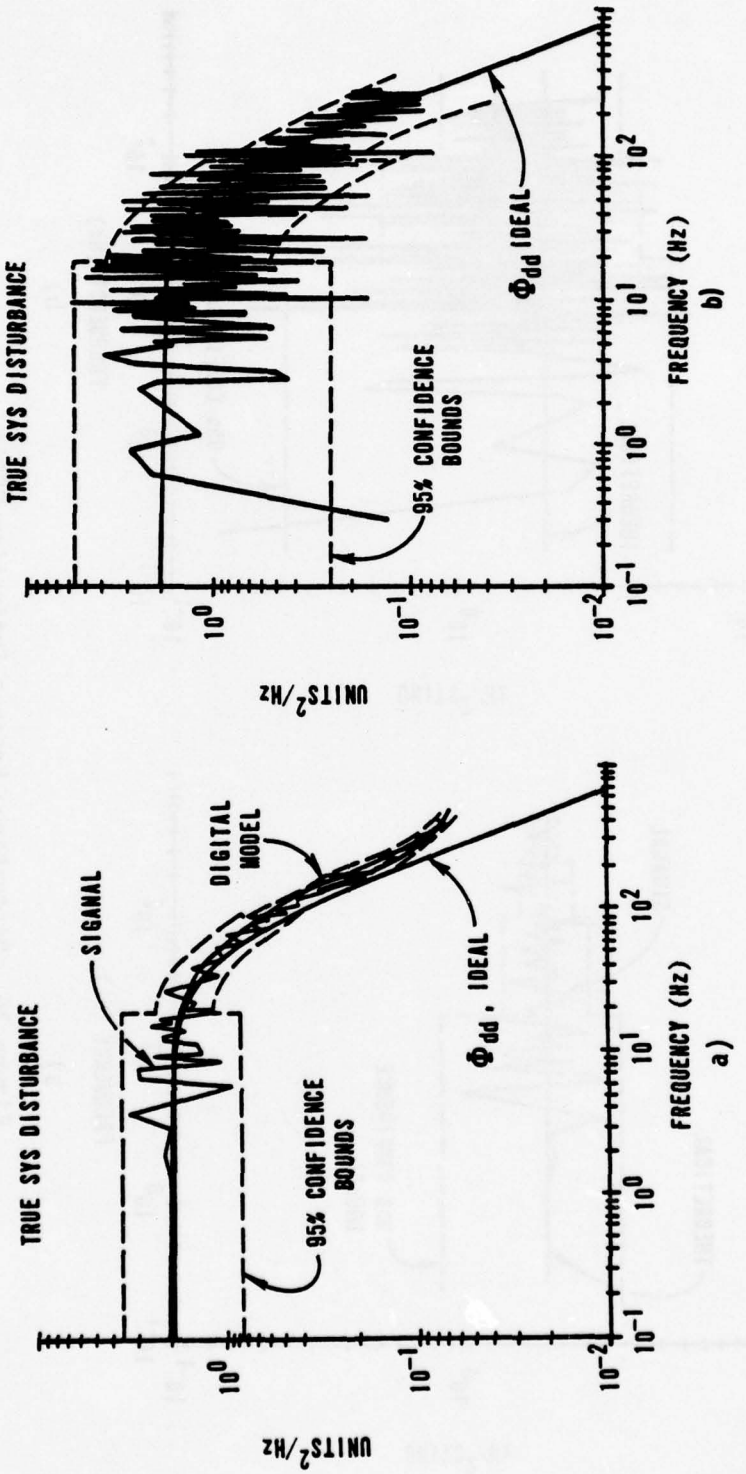


Figure 35. True Disturbance PSD.

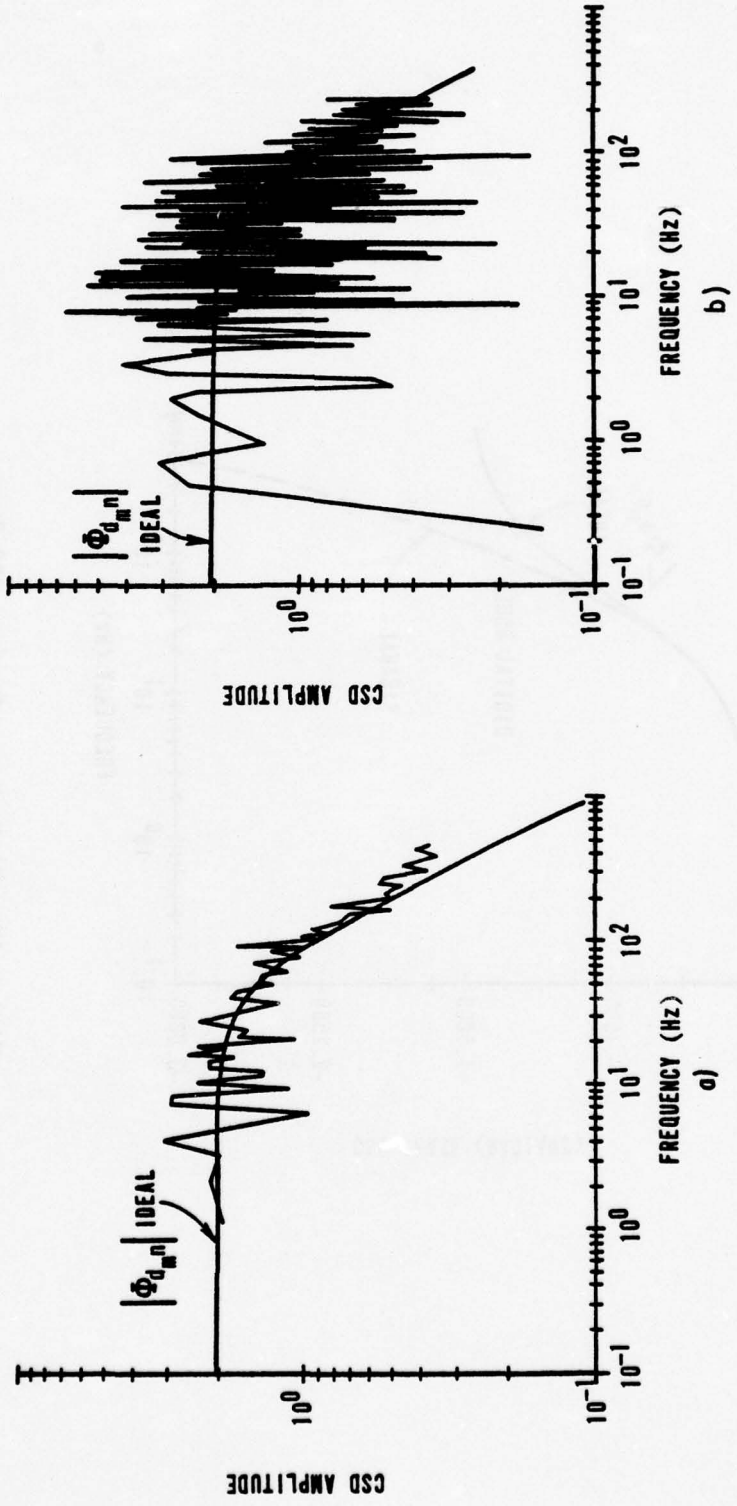


Figure 36. Disturbance Process Cross-Spectral Densities.

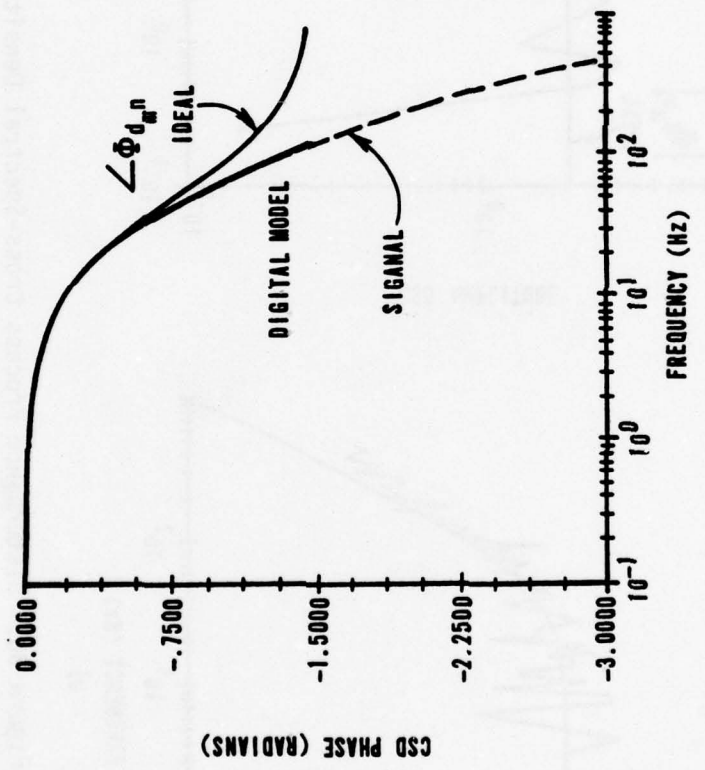


Figure 37. Disturbance Process CSD Phase.

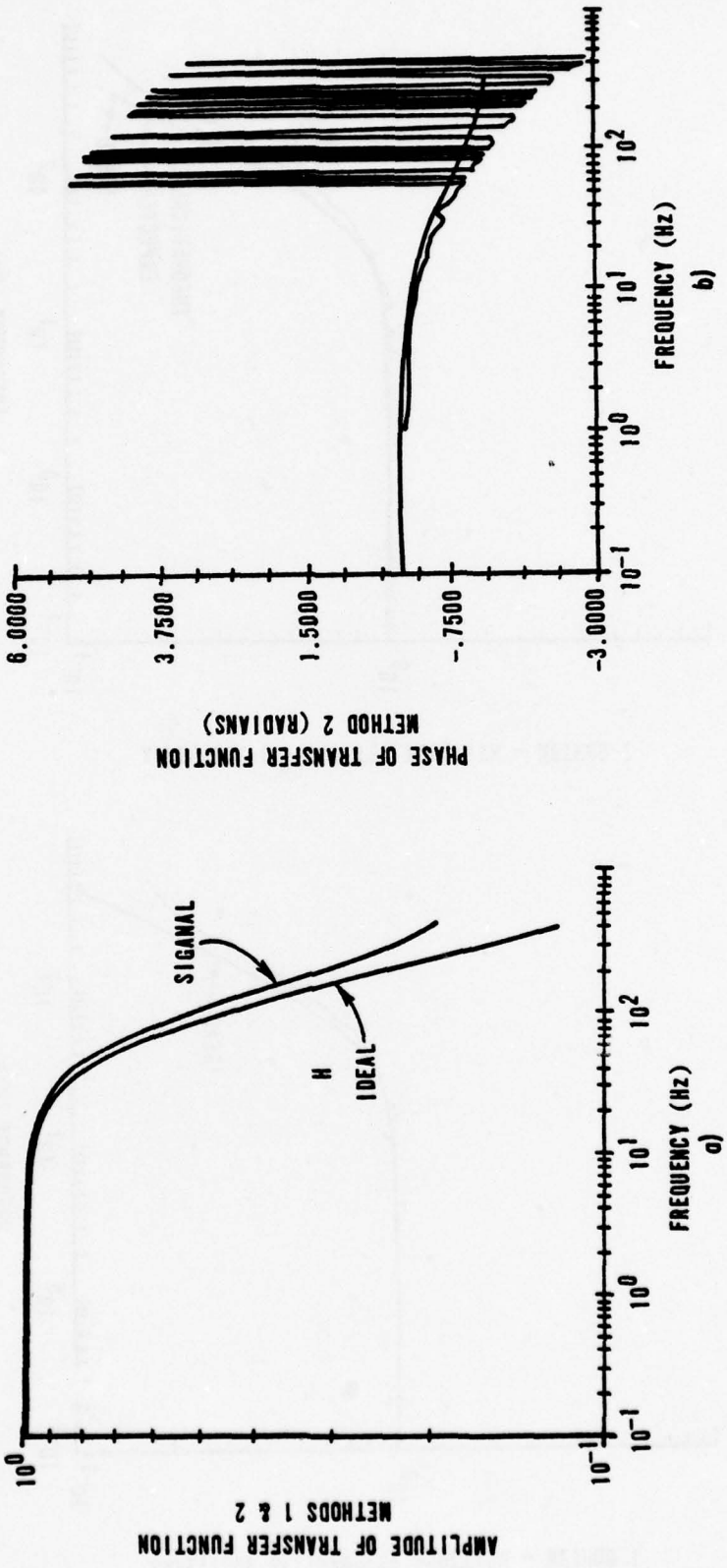


Figure 38. Disturbance Process Transfer Function Estimates, No Noise.

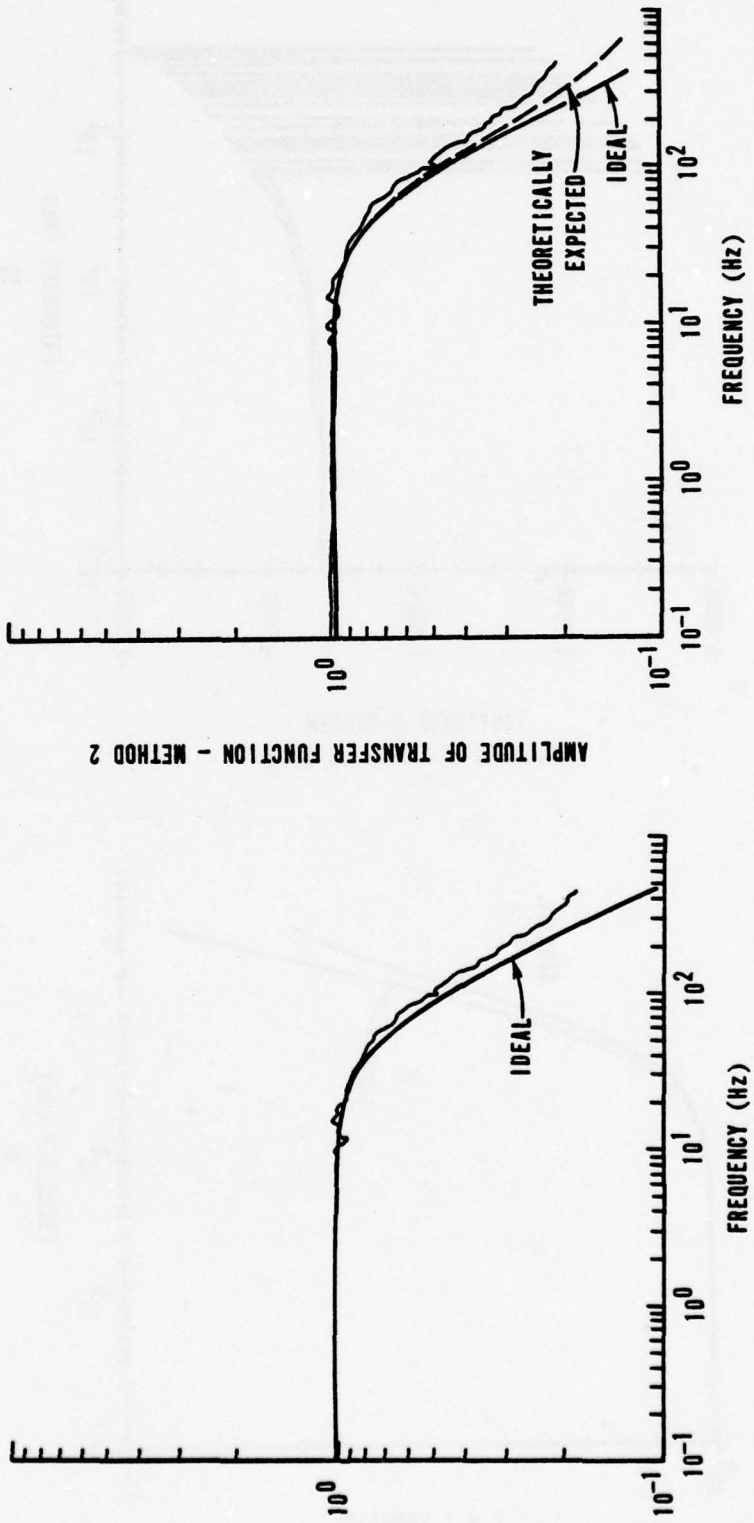


Figure 39. Disturbance Process Transfer Function Estimates, $\Phi_{ww} = .02$.

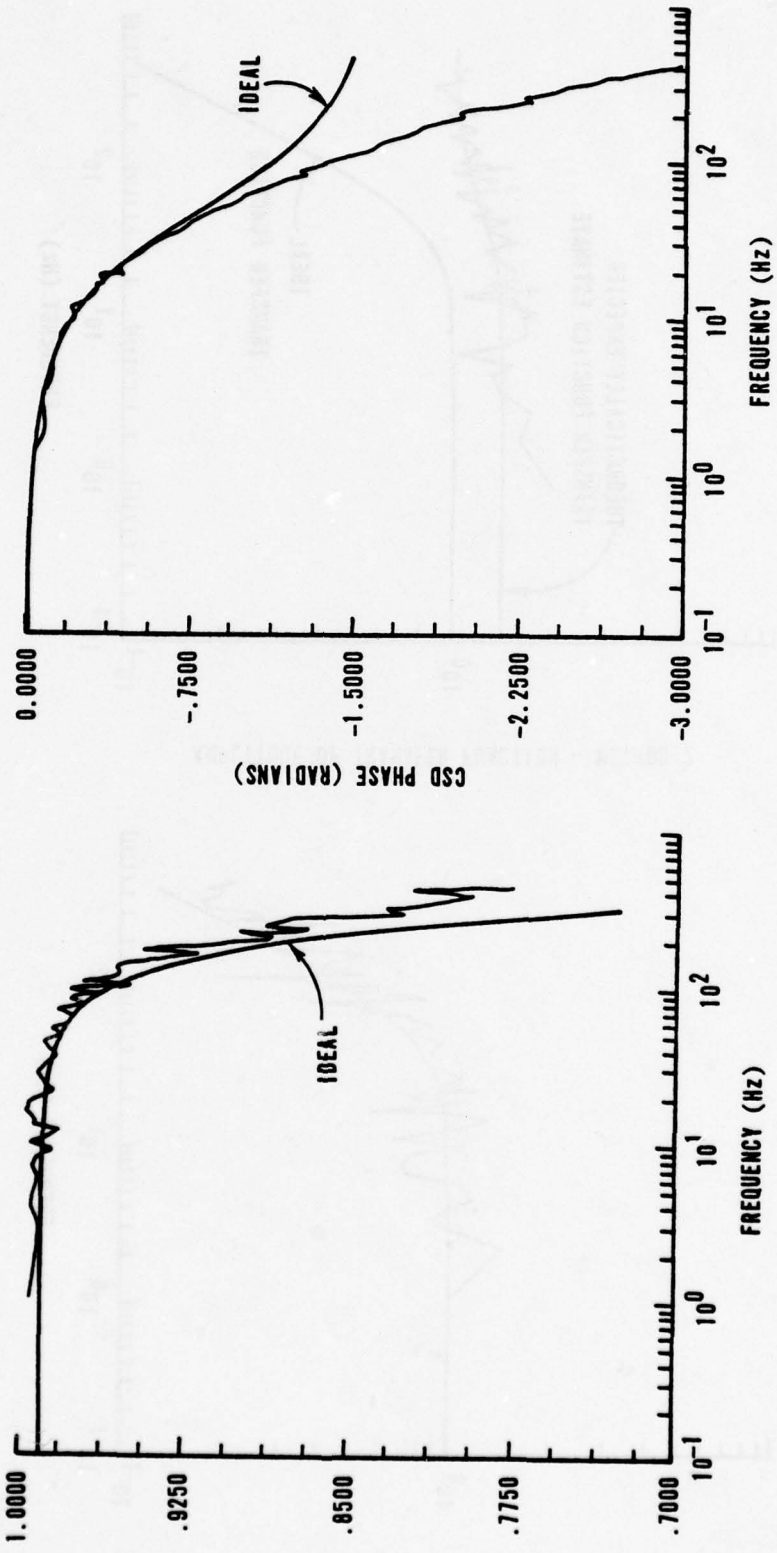


Figure 39 (Continued). Disturbance Process Transfer Function Estimates, Φ_{ww} = .02.

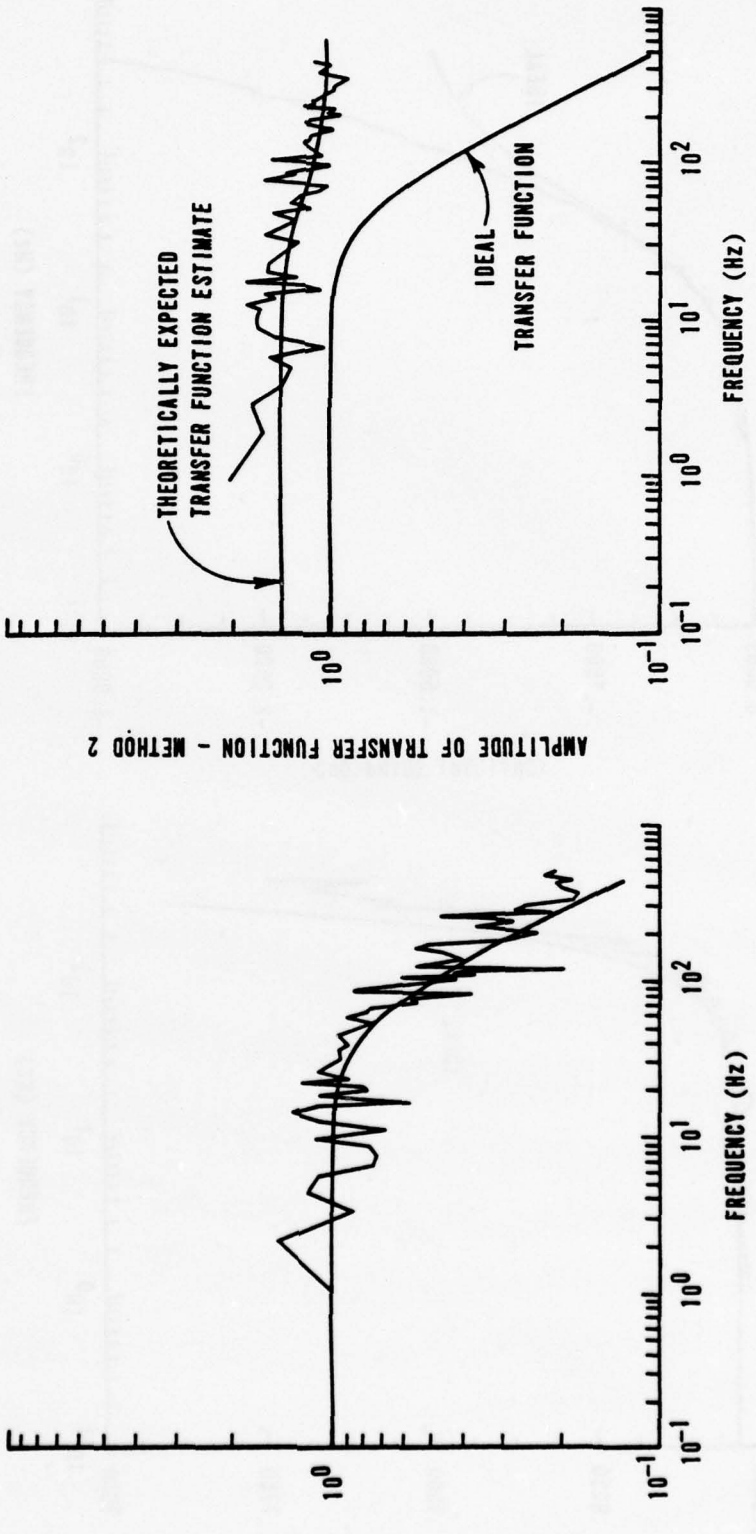


Figure 40. Disturbance Process Transfer Function Estimates, $\Phi_{ww} = 2.0$.

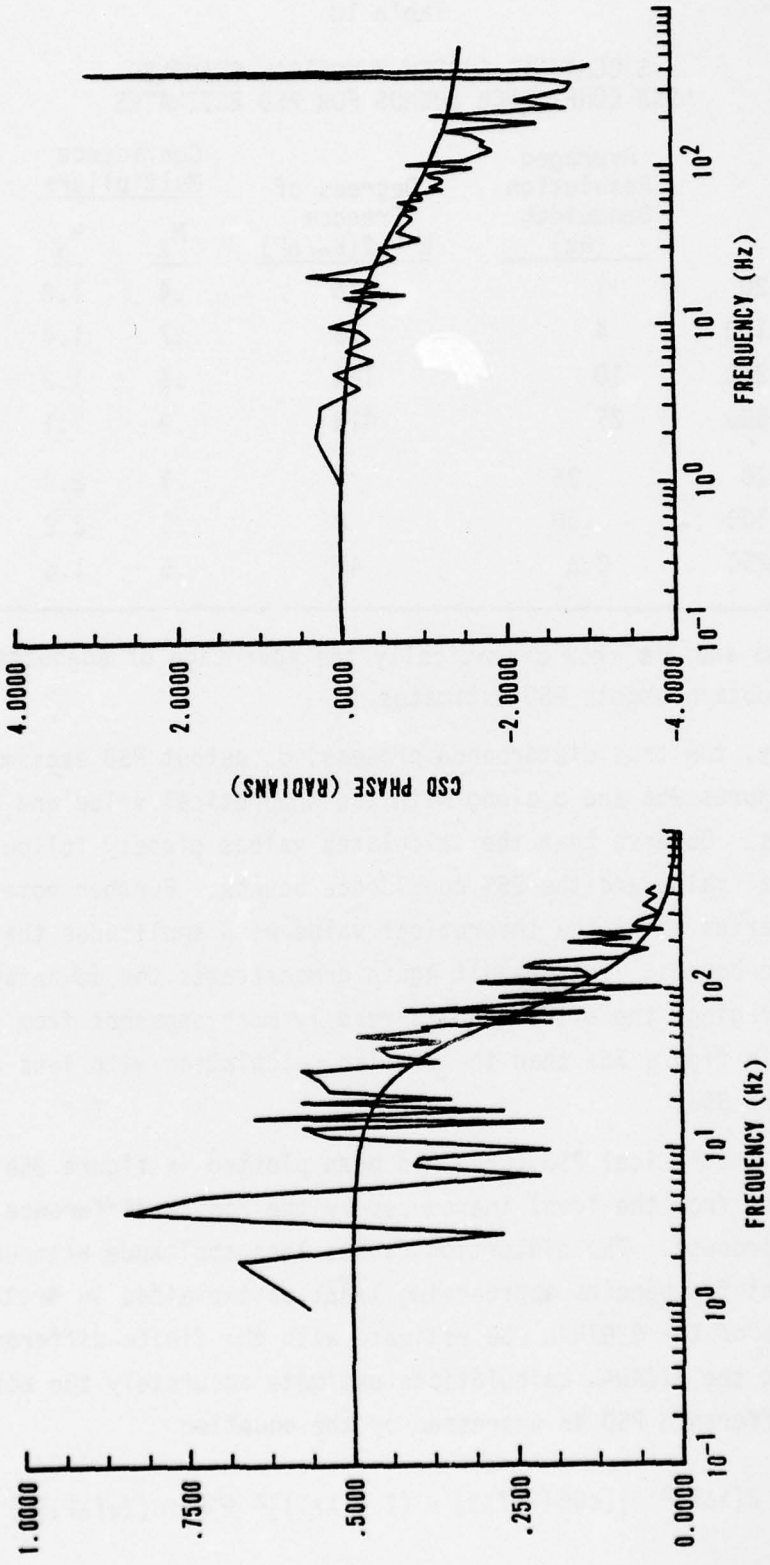


Figure 40 (Continued). Disturbance Process Transfer Function Estimates, Φ_{ww} = 2.0.

Table 10

STOCHASTIC SYSTEM NUMERICAL EXAMPLE
95% CONFIDENCE BOUNDS FOR PSD ESTIMATES

	Frequency Band (Hz)	Averaged Resolution Bandwidth (Hz)	Degrees of Freedom N = 2(BW/ΔF)	Confidence Multipliers M_ℓ	Confidence Multipliers M_μ	Spread (magnitude)	
Case A	0-20	1	16	.4	1.8	4.1	Case A
	20-100	4	65	.7	1.4	2.0	
	100-250	10	164	.8	1.2	1.6	
	250-500	25	410	.9	1.1	1.3	
Case B	0-20	.25	4	.1	2.8	23	Case B
	20-100	.50	8	.3	2.2	8.1	
	100-250	2.5	41	.6	1.6	2.4	

of figures 34a and 34b show dramatically the advantage of adequate frequency averaging to obtain smooth PSD estimates.

Similarly, the true disturbance process, d, output PSD estiamte ϕ_{dd} is plotted in figures 35a and b along with the theoretical value and the 95% confidence bounds. Observe that the calculated values closely follow the trend of the theoretical value and the 95% confidence bounds. Further note that the calculated PSD varies about the theoretical value with amplitudes that scale with the confidence bounds. This result again demonstrates the advantage of adequate frequency averaging; the actual PSD is readily more apparent from the analysis calculations in figure 35a than the similar calculation with less averaging shown in figure 35b.

A second theoretical PSD curve has been plotted in figure 35a to illustrate the distortions from the ideal introduced by the finite-difference equations used to model the process. The distortion causes less amplitude attenuation than the ideal filter at frequencies approaching $1/2\Delta T$ as explained in section IV.2. The correspondence of the SIGANAL PSD estimate with the finite-difference theoretical PSD shows that the SIGANAL calculations estimate accurately the actual signal PSD. The finite-difference PSD is expressed by the equation

$$\Phi = 2(\lambda\Delta T)^2 \left\{[\cos(2\pi f\Delta t) - (1 - \lambda\Delta T)]^2 + [\sin(2\pi f\Delta T)]^2\right\}^{-1} \qquad (4\text{-}31)$$

which is obtained from the finite difference equation, equation (4-7), by methods described in references 6, 10, 11, 17, and 18.

The disturbance process is a low-pass process. At frequencies below 10 Hz the white noise excitation is passed by the filter without attenuation (compare figures 34a and 34b). At frequencies above 10 Hz, the process attenuates the input by a factor that increases with frequency as described by equation (4-14). This example demonstrates the filtering action of linear systems.

b. Disturbance Process CSD's

Given the disturbance process excitation n and the output y, cross-spectral densities are computed. The CSD amplitude is simply the amplitude of the input spectra, in this case a constant $\Phi_{nn} = 2.0$, multiplied by the amplitude of the disturbance process transfer function. The SIGANAL CSD's are presented in figures 36a and b where different frequency averaging schedules have been used as in the previous figures and as listed in table 10. The theoretically expected CSD amplitude is followed by the SIGANAL calculations with only statistical variations due to the finite data observation interval. Like the PSD, the CSD estimate improves with more averaging. Exact confidence bounds for CSD's have not been calculated, but figure 36 suggests that they are approximately the same as PSD bounds.

The disturbance process input output cross-spectral density phase is plotted in figure 37. The same phase curve is obtained independently of the frequency averaging used. The SIGANAL CSD phase calculation deviates somewhat from the ideal phase curve but follows exactly the phase of the finite-difference numeric model. As in the case of the disturbance process output PSD, differences between the ideal and the calculated phase values stem from the distortions of the finite-difference simulation equations rather than from inaccuracies of the CSD phase calculation. For reference, the ideal phase ϕ is obtained from the process transfer function equation (4-13) as

$$\phi_{ideal} = -\tan^{-1}\left(\frac{2\pi f}{\lambda}\right) \tag{4-32}$$

while the phase associated with the finite-difference model is

$$\phi_{F-D} = -\tan^{-1}\left[\frac{\sin(2\pi f \Delta T)}{\cos(2\pi f \Delta T) - (1 - \lambda \Delta T)}\right] \tag{4-33}$$

The CSD phase calculation does not exhibit the statistical variation we saw in the PSD and CSD magnitude estimates. This result is typically valid whenever the signals used to generate the CSD estimate represent the actual input and outputs of a linear system and that further, neither measurement is corrupted by noise. Examples subsequently presented (e.g., figure 40d) show that the CSD phase signal does exhibit variations from the true linear system phase characteristic when one, or both, of the signals become corrupted by an additive noise signal.

c. Disturbance Process Transfer Function Estimates

Transfer function estimates of the disturbance process are obtained from the input and output PSD's and their CSD. The Method 1 transfer function amplitude is the CSD magnitude normalized by the input excitation PSD, Φ_{nn}. The resulting Method 1 transfer function estimate is plotted in figure 38a. The Method 2 estimate obtained by taking the square root of the ratio of the output PSD to the input PSD as in equation (4-21) matches exactly the Method 1 estimate since there are no extraneous noise sources present. Both estimates deviate from the ideal transfer function, also plotted, but match exactly the actually simulated transfer function. Figure 38a shows that precise transfer function estimates can be calculated by either Method 1 or Method 2 techniques.

The transfer function phase is simply the CSD phase in Method 1. No technique for calculating the transfer function phase is presented in the Method 2 analysis. Those familiar with linear systems theory might suggest that a phase estimate could be obtained by differencing the Fourier transform phases of the system output and input. The phase estimate obtained in this manner is plotted in figure 38b along with the ideal phase. The estimated phase is extremely noisy, especially at higher frequencies, and is attributable to the phase uncertainty of Fourier transforms of stochastic processes, even when there are no additive noise signals. Because of their high uncertainty, phase estimates obtained in this manner are seldom used in practice. We shall restrict our attention henceforth to CSD phase estimates.

The coherence function is not plotted for this no noise case. With no noise, the noise-to-signal ratio terms vanish [e.g., $\Phi_{ww} = 0$ in equation (4-22)] and the ideal coherence function is unity over the entire frequency band. The coherence calculated by the SIGANAL routines is not precisely unity because of numerical roundoff. However, the calculated coherence values are in the range (.998, 1] which indicates extremely high coherence between the process input and

output signals. We expect high coherence because the output signal is totally generated by application of the input signal to the disturbance process.

d. Noise Corrupted Disturbance Process

The previous example dealt with a noise free measurement of the process input and output signals. In the following, only the noise corrupted process output signal d_m is available for analysis. The additive noise signal w adds biases and variations to signal PSD and transfer function estimates. Transfer function estimates by Methods 1 and 2, coherence calculations and the CSD phase are plotted in figures 39a - d. In this case, Φ_{ww} equals .02 corresponding to a noise to signal (N/S) ratio of 1/100 in the low frequency band (< 50 Hz). Note that the transfer function magnitude estimates display a more erratic character than the noise free estimates in figure 38. The presence of corruptive noise is indicated by the coherence function which is biased from unity and which diverges further from unity at frequencies beyond 50 Hz. This behavior is predicted by a N/S analysis. The constant amplitude noise Φ_{ww} = .02 becomes a larger fraction of the signal PSD Φ_{dd} which is increasingly attenuated at higher frequencies.

The degrading effect of higher N/S ratios is demonstrated in figure 40. A low frequency band N/S ratio of 1 is simulated by increasing Φ_{ww} to 2.0. Both Method 1 and Method 2 transfer function amplitude estimates and the CSD phase become appreciably erratic. The Method 2 transfer function estimate has also become biased from the true value. This bias arises from the output measurement noise which adds a bias to the true process output PSD as shown in equation (4-16). The biased output PSD biases the transfer function estimate. The presence of additive noise components corresponding to large N/S ratios is indicated by the coherence function, figure 40c which has dropped to at most .5. As shown in section III, a coherence of .5 shows that only half of the output signal power is attributable to a linear system excited by the input signal. Thus, as inspection of figure 30a shows, Φ_{ww} contributes to the measured output power; Φ_{ww} contributes one-half the total output power in this case.

e. Control System PSD's

The control system analysis results are presented in figures 41-44. A summary of the theoretical transfer function estimates, which are well-matched by the SIGANAL results, are plotted in figures 45-47.

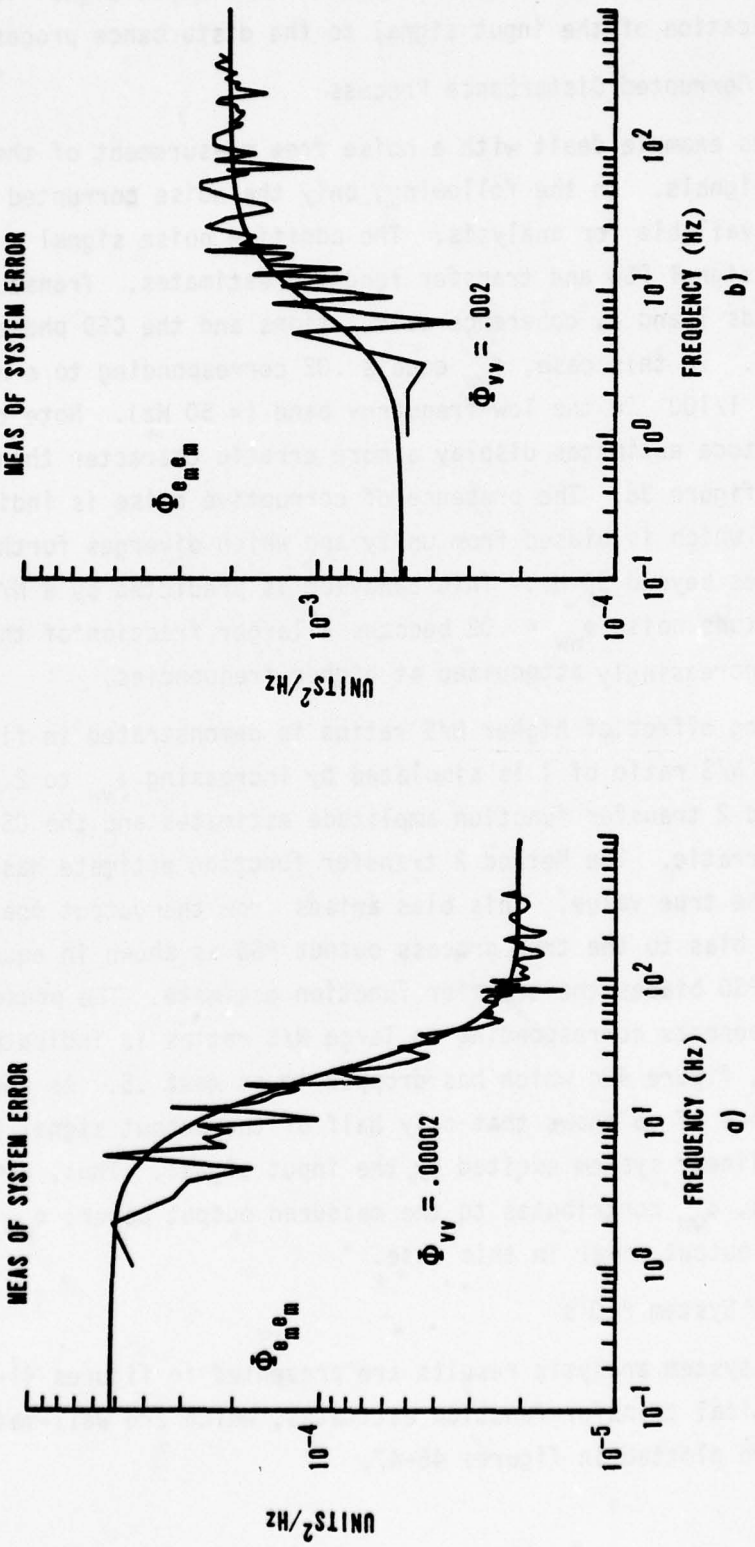


Figure 41. Control System Error PSD's.

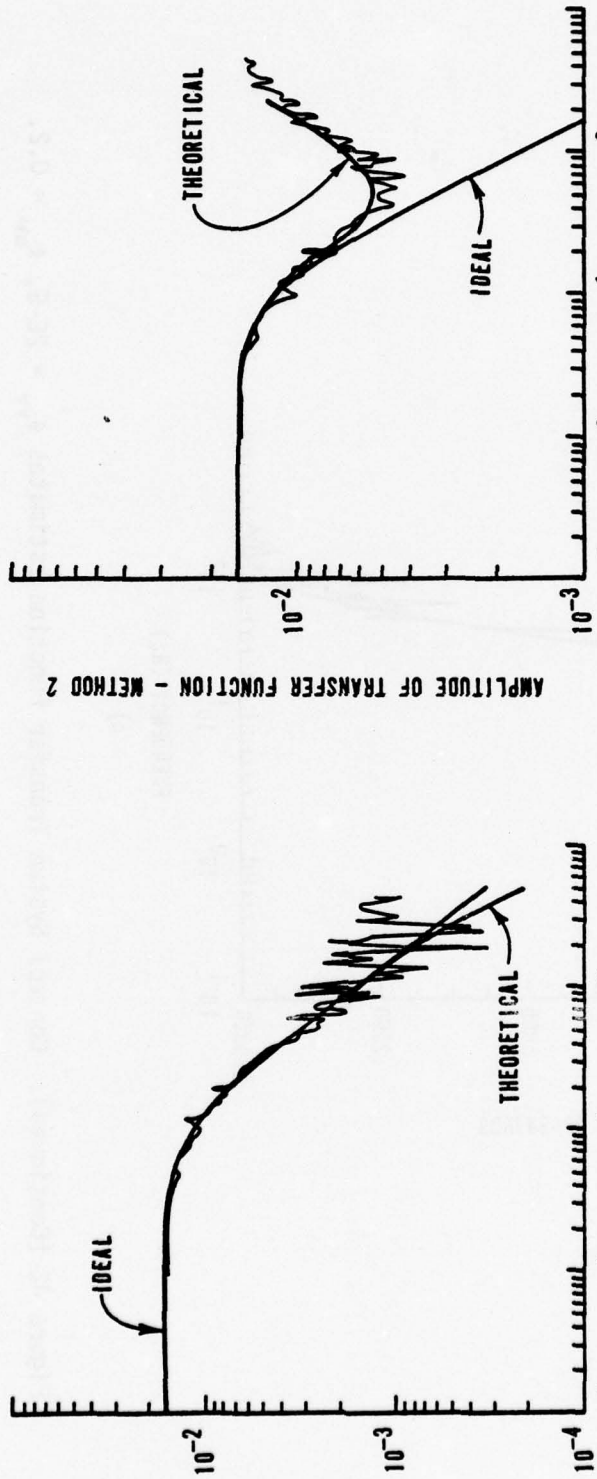


Figure 42. Control System Transfer Function Estimates Φ_{vv} = 2E-5, Φ_{ww} = .02.

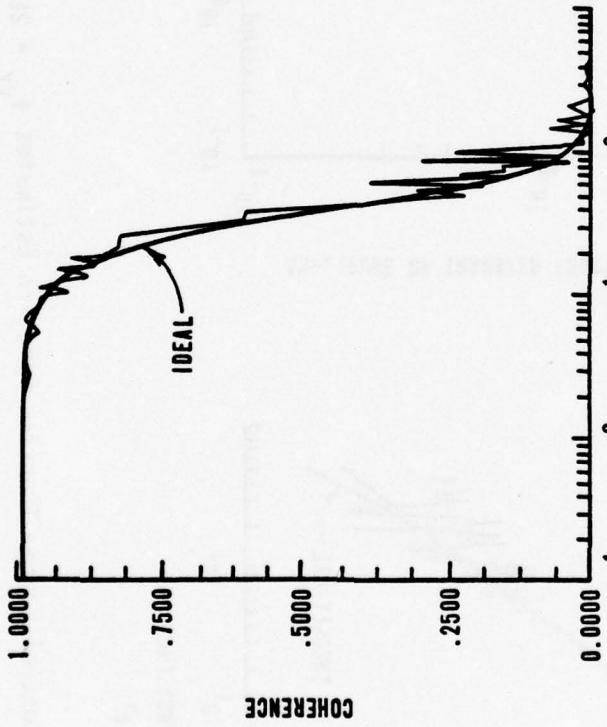


Figure 42 (Continued). Control System Transfer Function Estimates Φ_{vv} = 2E-5, Φ_{ww} = 0.2.

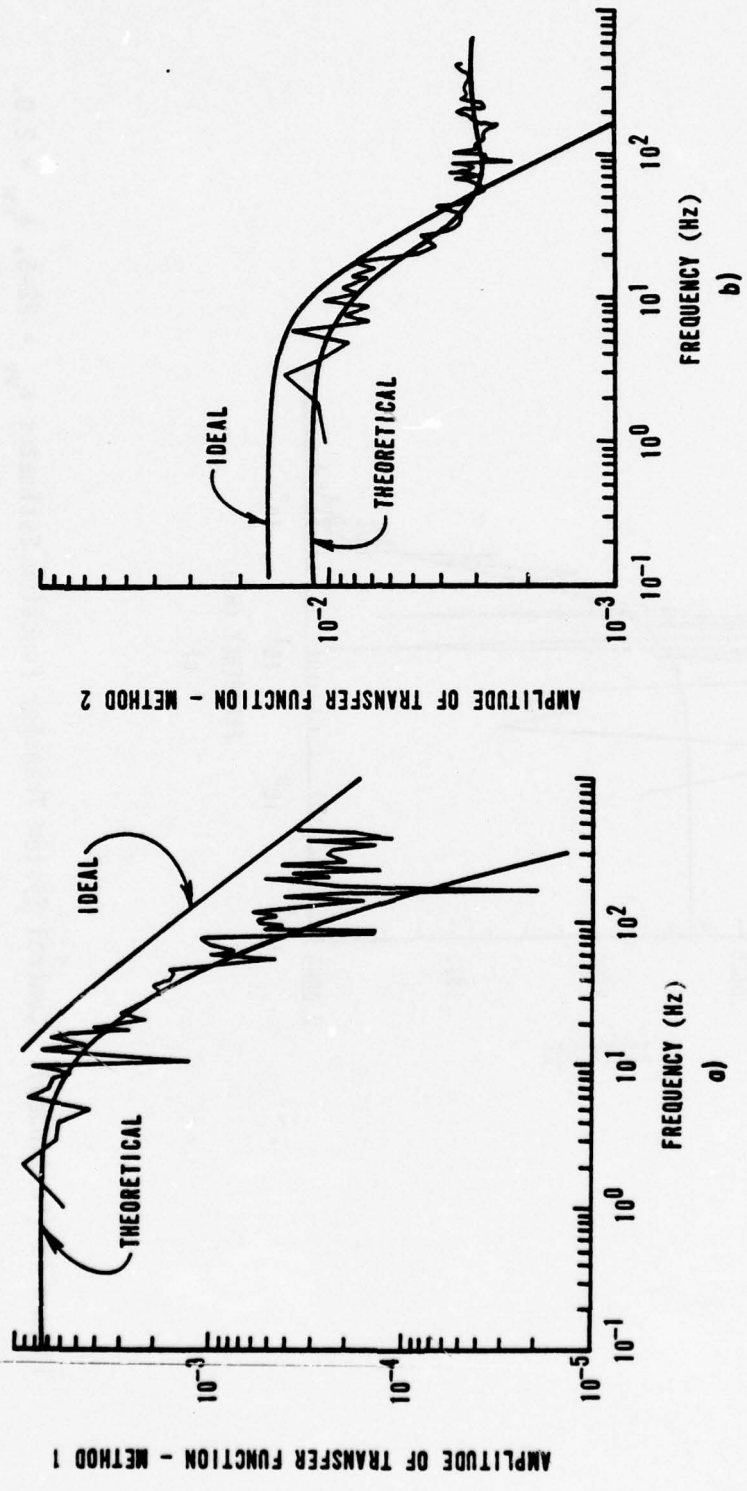


Figure 43. Control System Transfer Function Estimates Φ_{vv} = 2E-5, Φ_{ww} = 2.0.

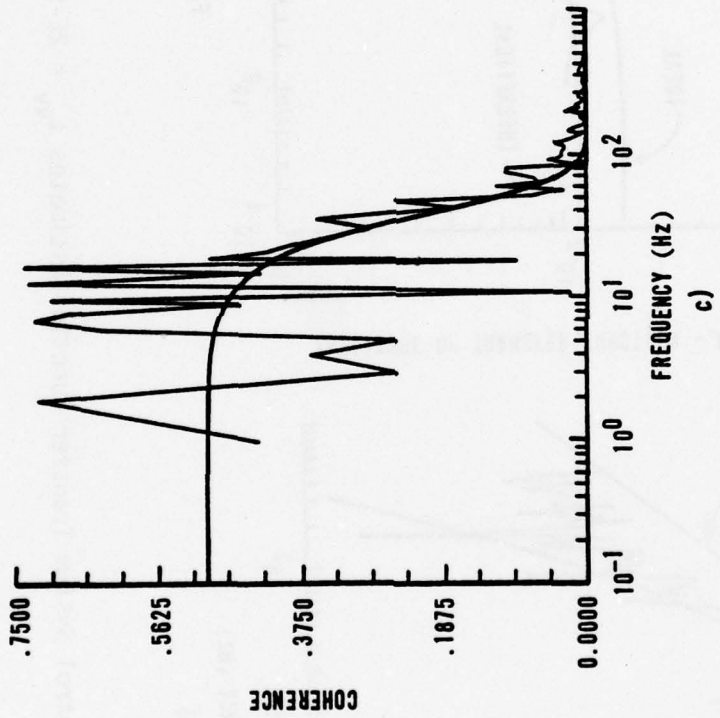


Figure 43 (Continued). Control System Transfer Function Estimates Φ_{vv} = 2E-5, Φ_{ww} = 2.0.

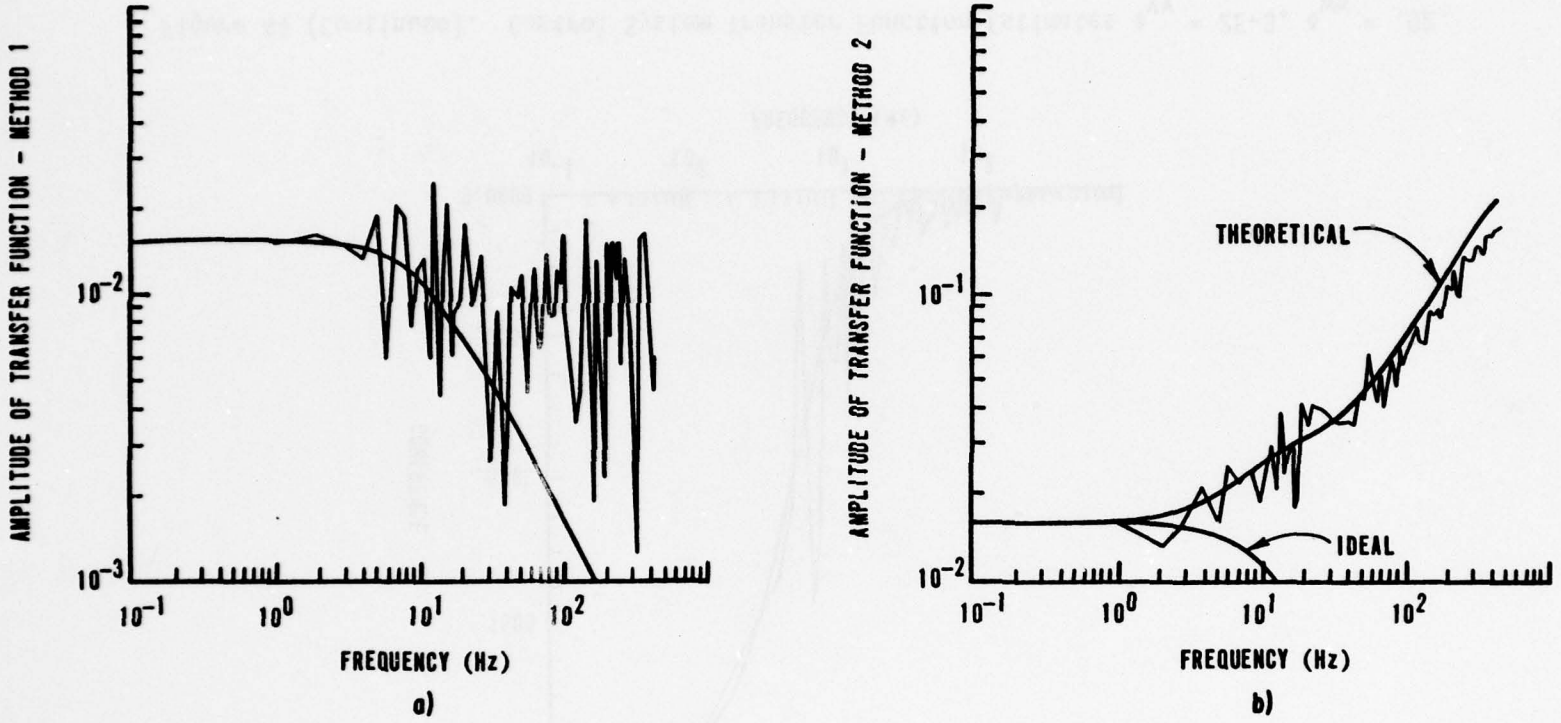


Figure 44. Control System Transfer Function Estimates Φ_{VV} = 2E-3, Φ_{WW} = .02.

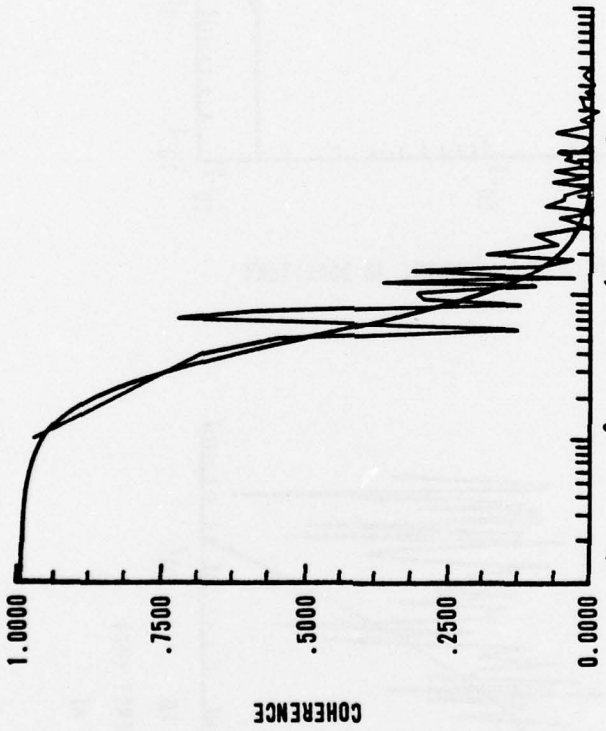


Figure 44 (Continued). Control System Transfer Function Estimates Φ_{vv} = 2E-3, Φ_{ww} = .02.

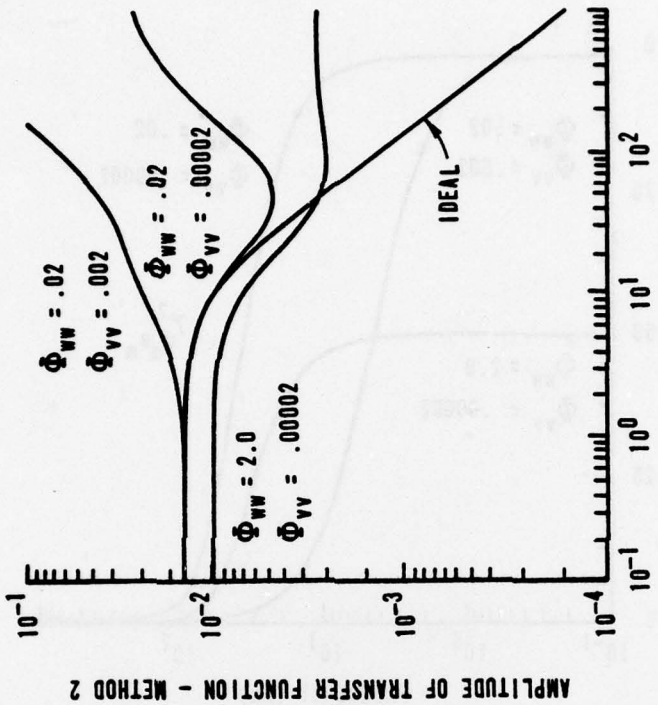


Figure 46. Method 2 Transfer Function Estimate Summary.

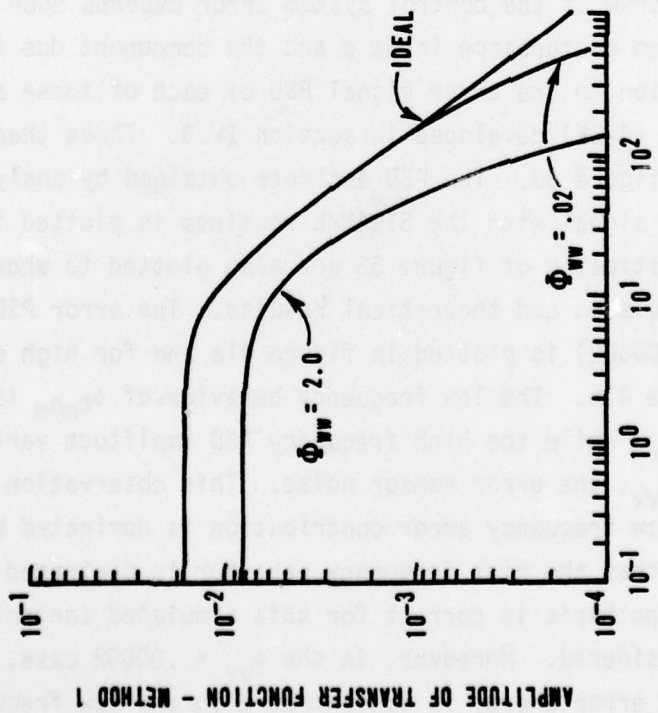


Figure 45. Method 1 Transfer Function Estimate Summary.

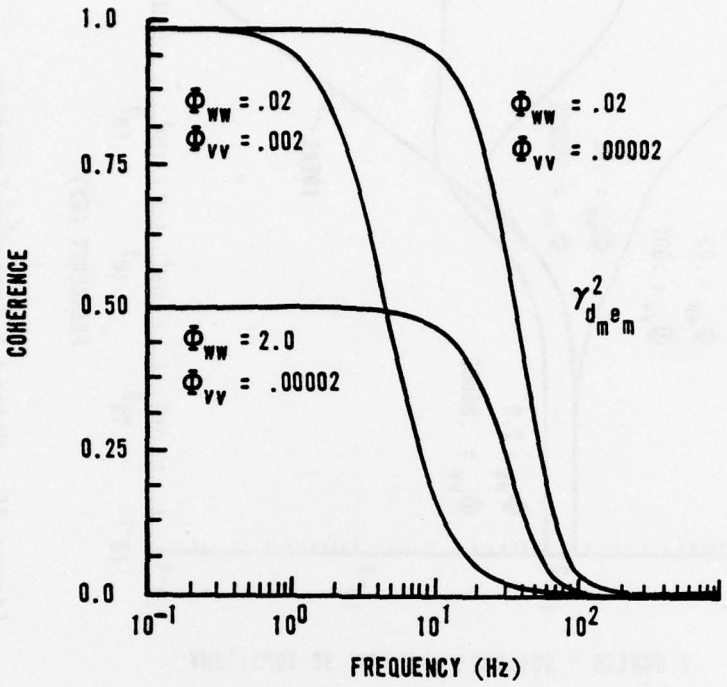


Figure 47. Coherence Function Summary.

The power spectrum of the control system error depends upon the error perturbation of the system disturbance input d and the component due to the sensor error. The contribution to the error signal PSD of each of these sources is characterized by equation (4-25) developed in section IV.3. These theoretical results are plotted in figure 33. The PSD estimate obtained by analyzing the digitally simulated error signal with the SIGANAL routines is plotted in figure 41. The theoretical PSD estimates of figure 33 are also plotted to show the excellent agreement of the calculated and theoretical results. The error PSD $\Phi_{e_m e_m}$ for low sensor noise (Φ_{vv} = .00002) is plotted in figure 41a and for high sensor noise (Φ_{vv} = .002) in figure 41b. The low frequency behavior of $\Phi_{e_m e_m}$ is unalterated by these variations in Φ_{vv} while the high frequency PSD amplitude varies directly with the changes in Φ_{vv}, the error sensor noise. This observation suggests the hypothesis that the low frequency error contribution is dominated by the disturbance process and that the high frequency behavior is dominated by the error sensor noise. The hypothesis is correct for this simulated control system and the ranges of Φ_{vv} considered. Moreover, in the Φ_{vv} = .00002 case, the error PSD shows that the system error energy is concentrated in the low frequencies (< 20 Hz

or so). When ϕ_{vv} is increased to .002, the control system error becomes concentrated in the high frequency band (> 20 Hz). Comparisons of figures 41a and 41b show that the RMS of the error is much larger in the latter, since, by equation (2-111), the RMS is given by the square root of the integral of $\Phi_{e_m e_m}$ over the entire frequency band, and since the PSD in figure 41b is greater in magnitude than the PSD of figure 41a.

f. Control System Transfer Function Estimates

Estimates of the closed loop transfer function from the disturbance excitation (input) to the system error (output) are obtained by appropriately processing measurements of the system disturbance and the system error. These estimates are obtained by using Methods 1 and 2 transfer function formulae as had been used to obtain disturbance process transfer function estimates. The control process transfer function estimate problem adds several additional features not exhibited in the disturbance process example. They are:

- The system input signal measurement, as well as the output, is corrupted by additive measurement noise.
- The system input signal PSD varies with frequency (the disturbance is a "colored" or correlated process).
- The output measurement noise is influenced by the control system in which the noisy error (output) sensor is embedded.

The transfer function estimates are presented in figures 42-44 for three cases of noise level. In figure 42 the case of low measurement noise is examined. The high input measurement noise case (with low output noise) is examined in figure 43. High output noise but low input noise is considered in figure 44. Method 1 transfer function SIGANAL results are plotted in part a of each figure, Method 2 results are plotted in part b and the coherence function results are plotted in part c. Also plotted in each figure are the theoretical results that we would expect to obtain given perfect knowledge of the system transfer functions, the true disturbance PSD and the measurement noise PSDs. These theoretical results are plots of equations (4-28), (4-29), and (4-30) derived in section IV.3. These theoretical results are summarized in figures 45-47. Also plotted with the transfer function estimates is the ideal (or actual) transfer function, which typically differs from the theoretical transfer function. Differences between the ideal and the theoretical transfer functions are distortions introduced into the theoretical transfer function estimate by the noise signals inherent in the signal measurements. These distortions are systematic errors and cannot be removed from the transfer function estimation process.

Except for the case of the high error sensor noise Method 1 transfer function estimate in figure 44a, the SIGANAL transfer function results, obtained by processing the simulated signals, agree very closely with the theoretically expected results. Thus, the SIGANAL processing is shown to provide, in a practical data analysis situation, accurate results. The erratic variations of the computed estimates about the theoretical curves are variations within the statistical confidence as limited by the finite duration data interval processed. These variations would increase if less frequency averaging were employed to smooth the raw estimates. Conversely, the variations could be reduced by frequency averaging more adjacent estimates to increase the number of degrees of freedom. The estimates could also be improved by processing a longer data interval. The Case A frequency averaging schedule listed in table 10 is employed to obtain the estimates plotted in figures 41-44.

The anomalous variation of SIGANAL and theoretical results in figure 44a stems from a low statistical confidence of the SIGANAL result in the frequency band of 30 Hz and above. Deviation of statistical confidence bounds for transfer function estimates is extremely complex, is data (transfer function) dependent, and often cannot be practically and accurately calculated in a given problem. Nonetheless, we argue that the variation of results exists because of low statistical confidence caused by the relatively high noise level of the error signal in the high frequency band. Reference figure 41b. The cross-spectral density between the input and output (error) signals in this high frequency band, while theoretically having near zero magnitude, computationally has components large compared with with the true CSD, unless enough averaging can be performed to improve the computational average. Arguments as to the source of this variation are academic in the practical data analysis case where we must ignore or disregard transfer function estimates over frequency bands in which the coherence is low. In this case, the coherence is low above 5 Hz and we must ignore the transfer function estimates.

The coherence function indicates the relative level of noise present in either the input or output measurement. The coherence function decreases as the noise level increases. Since increased noise levels bias and corrupt our transfer function estimates, we must discount these estimates in high noise situations, which we can determine from the coherence function. As a rule of thumb, we should not trust estimates made in frequency bands for which the coherence is 0.5 or less. Theoretically, output signal noise does not effect the Method 1 transfer function estimate. Unfortunately, there is no technique by which we can take advantage of

this fact since we cannot determine whether the input or the output signal noise is causing low coherence.

Figures 45 and 46 show that the Method 1 transfer function estimate is the better estimate. This result is generally true since Method 1 estimates are sensitive to only input signal measurement noise while Method 2 estimates are sensitive to both input and output measurement noises. Examination of the coherence functions plotted in figure 47 shows that the best transfer function estimates are obtained at low frequencies. At higher frequencies, the N/S increase thus lowering the coherence and corrupting the transfer function estimates.

g. Miscellaneous Topics

The simulated signal analysis example provides a mechanism for additional insight and analysis that might be applied. The character of stochastic process signals is illustrated in figures 48a and 48b in which the time histories of two simulated signals are plotted. The control system output signal y is plotted in figure 48a. This signal is limited in frequency content to 10 Hz by the bandwidth of the control system. Notice the slow temporal variations of the signal. The true disturbance signal plotted in figure 48b has a noticeably higher frequency content. This signal is the output of a 50 Hz process which does contain higher frequency components. A review of the signal time histories, as these plots provide, should be performed in any analysis to determine whether the signal behavior is random, as the figure 48 plots appear, or whether there are deterministic components characterized by trends, discontinuities, or low frequency variations.

Another useful analysis is a probability density function calculation of the measured signal values. The probability density function indicates the relative frequency with which signal values occur. The pdf also characterizes signal mean values and the dispersion or variance of the signal values about the mean. Finally, the pdf allows the analyst to determine if the density function is gaussian, symmetric, or skewed and whether or not there are anomalous data points.

Probability density calculations of the white noise disturbance process excitation signal n are plotted in figure 49. In figure 49a the pdf of the n is compared with a theoretical gaussian probability density function parameterized by the signal mean and variance. The excellent correspondence indicates that n is derived from a gaussian process. The random number generator from which n is directly obtained is designed to produce gaussian deviates. The pdf plotted in figure 49b is the same signal n after it has been processed by the SIGANAL TAPDC

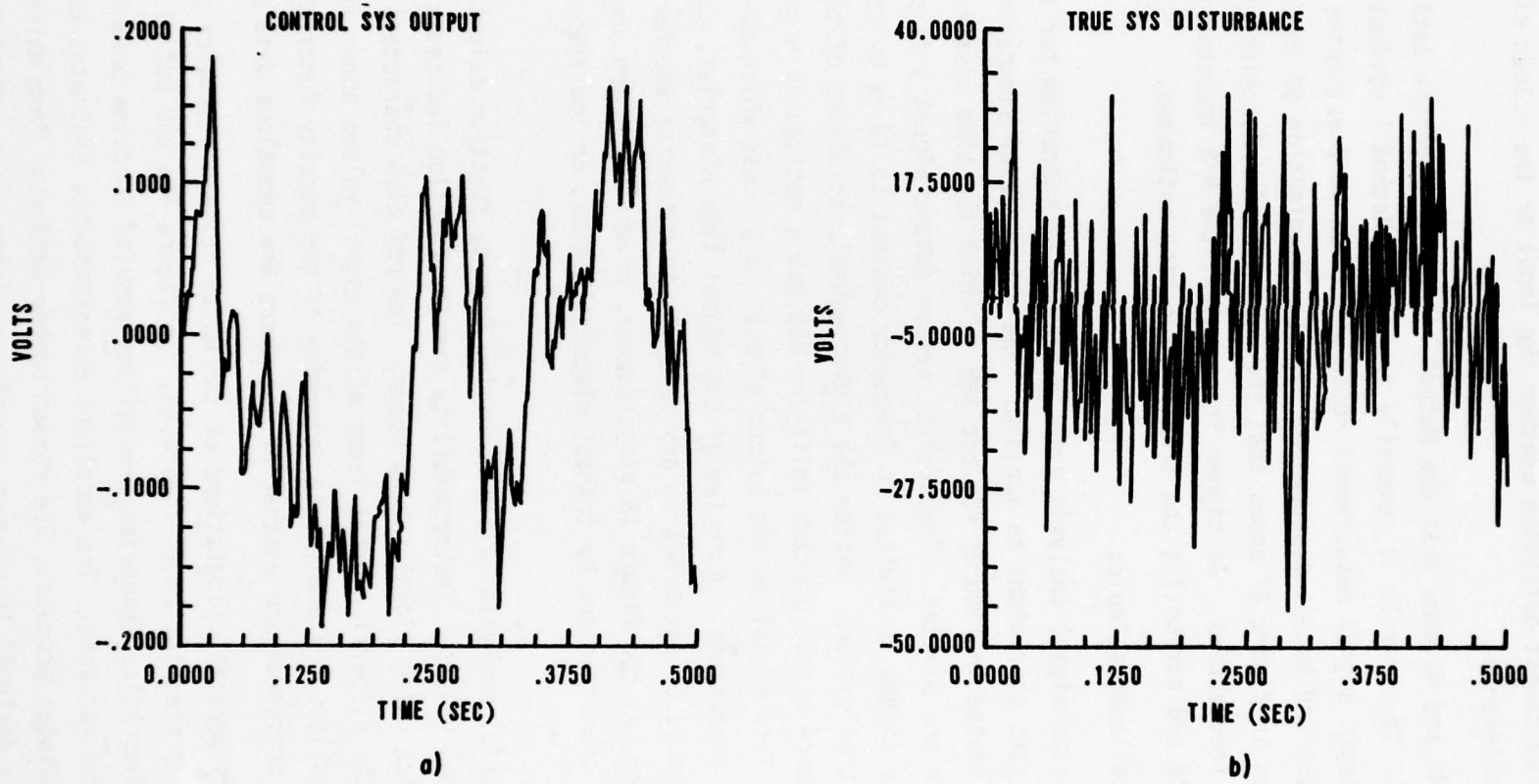


Figure 48. Typical Signal Time History.

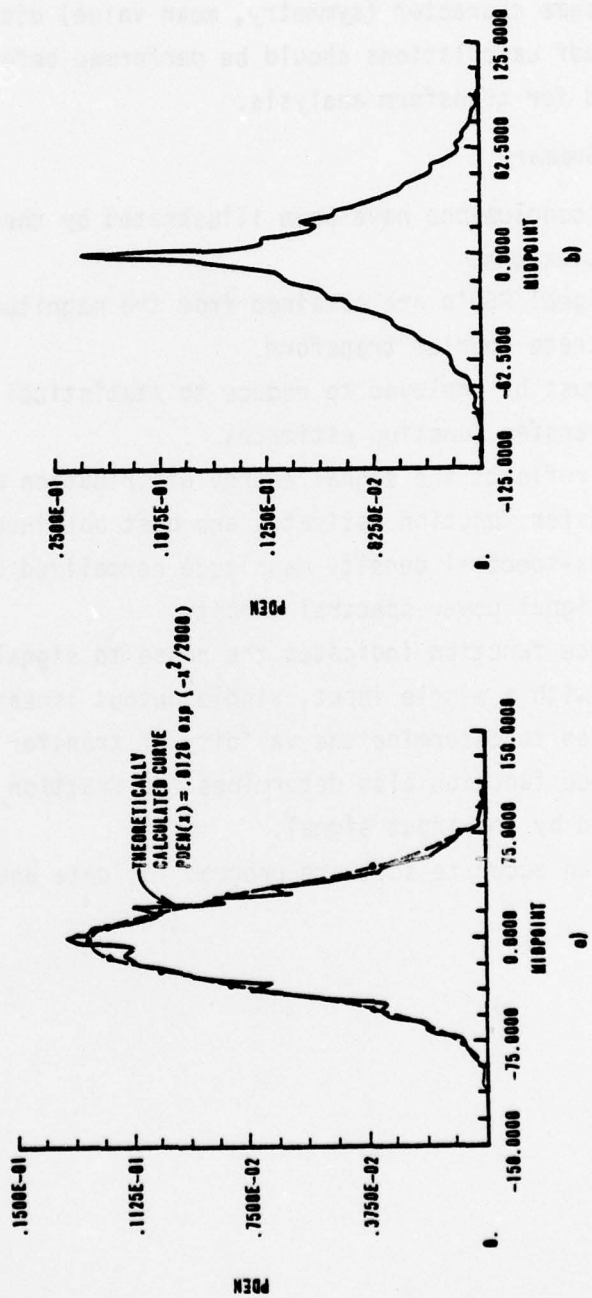


Figure 49. Signal Probability Density Function Estimate.

routine. This program module removes signal biases and trends and multiplies the result by the cosine data window described in Section III. Multiplication of the data by the window function alters the pdf of the original signal by adding a concentration of new values near zero; hence the sharp peak of the pdf. Otherwise the pdf retains the same character (symmetry, mean value) displayed in figure 49a. Generally, however, pdf calculations should be performed before the TAPDC signal modification required for transform analysis.

h. Example Summary

The following conclusions have been illustrated by the analysis example and the theoretical calculations.

- Accurate signal PSD's are obtained from the magnitude squared of the signal discrete Fourier transform.
- Averaging must be employed to reduce to statistical variation of PSD, CSD, and transfer function estimates.
- Signal PSD reflects the signal energy distribution with frequency.
- System transfer function estimates are best obtained as the input-output signal cross-spectral density magnitude normalized by the magnitude of the input signal power spectral density.
- The coherence function indicates the noise to signal levels of signals associated with a single input, single output linear system. It should be calculated to determine the validity of transfer function estimates. The coherence function also determines the fraction of output signal power caused by the input signal.
- SIGANAL is an accurate software program for data analysis of stochastic processes.

REFERENCES

1. Tenorio, R. A., SIGANAL: A Modular Signal Analysis Program, AFWL-TR-75-165, Air Force Weapons Laboratory, Kirtland Air Force Base, New Mexico, February 1976.

2. Bendat, J. S. and Piersol, A. G., Random Data: Analysis and Measurement Procedures, Wiley, 1971.

3. Jazwinski, A. H., Stochastic Processes and Filtering Theory, Academic Press, 1970.

4. Papoulis, A., Probability, Random Variables and Stochastic Processes, McGraw-Hill, 1965.

5. Sage, A. P. and Melsa, J. L., Estimation Theory With Applications to Communications and Control, McGraw-Hill, 1971.

6. DeRusso, P. M., et al., State Variables for Engineers, Wiley, 1965.

7. Wiener, N., Extrapolation, Interpolation, and Smoothing of Stationary Time Series, MIT Press, Cambridge, Massachusetts, 1966.

8. Koenigsberg, W. D., Spectral Analysis of Random Signals - Techniques and Interpretation, C. S. Draper Laboratory Report E-2771, June 1973.

9. Cooley, J. W. and Tukey, J. W., "An Algorithm for the Machine Calculation of Complex Fourier Series," Math, Comput., Vol. 19, pp. 297-301, April 1965.

10. Gold, B. and Rader, C., Digital Processing of Signals, McGraw-Hill, 1969.

11. Oppenheim, A. V. and Schaefer, R. W., Digital Signal Processing, Prentice-Hall, 1975.

12. Schwartz, M. and Shaw, L., Signal Processing: Discrete Spectral Analysis, Detection, and Estimation, McGraw-Hill, 1975.

13. Abramowitz, M. and Stegun, I. A., Handbook of Mathematical Functions, pp. 984-985, Dover Publications.

14. Blackman, R. B. and Tukey, J. W., The Measurement of Power Spectra, Dover Publications, 1958.

15. Halvorsen, W. G. and Bendat, J. S., "Noise Source Identification Using Coherent Output Power Spectra, Sound and Vibration, August 1975.

16. Bucy, R. S., Hecht, C., and Senne, K. D., An Engineers Guide to Building Nonlinear Filters, F. J. Seiler Laboratory TR-72-0004, Air Force Systems Command, May 1972.

17. Negro, J. E., Digital Algorithms for Estimating the Spectral Density of Deterministic Signals, M.S.E.E. Thesis, University of Wisconsin, June 1969.

18. Rabiner, L. R. and Gold, B., Theory and Application of Digital Signal Processing, Prentice-Hall, 1975.

NOMENCLATURE

A	sinusoid amplitude
C_{11}	covariance $\mathcal{E}[(y - \mu_y)\ (z - \mu_z)]$
$C_{k\ell}$	joint central moment $\mathcal{E}[(y - \mu_y)^k\ (z - \mu_z)^\ell]$
$C_{xy}(t_1, t_2)$	stochastic process covariance function
$C_{xy}(\tau)$	time-average covariance function
CF	Correction factor for power loss by data windowing
DFT	Discrete Fourier Transform
$\delta(t)$	Dirac delta function
$\delta_T(t)$	periodic train of delta functions - period ΔT
$\mathcal{E}[\cdot]$	expectation operator
e_d	error probability confidence bound
f	cycles per second frequency variable, $f=w/2\pi$
$F[\cdot]$	Fourier transform of $[\cdot]$
$f(y)$	probability density function - pdf
$F(y)$	probability distribution function
$f(x_1,t_1)$	stochastic process first-order density function
$f(x_1,x_2, t_1,t_2)$	stochastic process second-order density function
$F(x_1,x_2,x_n, t_1,t_2,t_n)$	stochastic process n^{th} order probability distribution function
FFT	Fast Fourier Transform
$G_{xx}(w)$	two-sided power spectral density - PSD
$G_{xy}(w)$	two-sided cross-spectral density - CSD
γ^2_{xy}	coherence function
$h(t)$	linear dynamic system impulse response
$H(w)$	linear system transfer function, $H(w) = F[h(t)]$
j	unit imaginary number, $j = \sqrt{-1}$
$m_{k\ell}$	joint moment, $\mathcal{E}[y^k z^\ell]$

$m_{xy}(t_1,t_2)$	stochastic process correlation function
M_ℓ	lower bound confidence multiplier for PSD's
M_u	upper bound confidence multiplier for PSD's
μ_y	mean value (average) of the random variable y
$\hat{\mu}$	mean value estimate
N	number of data points
P	time average signal power
P (•)	probability of the indicated (•) set, outcome or event
$P(x, f_o, \Delta f)$	average power density of signal x at frequency f_o with bandwidth Δf
Φ	phase variable
$\Phi_{xx}(f)$	one-sided power spectral density, PSD
$\Phi_{xy}(f)$	one-sided cross spectral density, CSD
$R_{xy}(\tau)$	time-average correlation
σ_y^2 (σ_y)	variance (standard deviation) of the random variable y
t	time variable
T	data observation interval
T_d	delay time
ΔT	sampling interval
w	radian frequency variable
x(t)	time function or signal
x_i	the i^{th} sample of the signal x
$\bar{x}_i$	time-average of the i^{th} ensemble element
$x_T(t)$	truncated time function
X(w)	Fourier transfer of x(t)
$X_p(k)$	discrete Fourier transform of x at the k^{th} frequency
{x}	a set or event